The influence of musical training on spectro-temporal speech processing

Jürg Kühnis

Influence of musical training on spectro-temporal speech processing

Thesis
Presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the degree of Doctor of Philosophy

by

Jürg Kühnis

Accepted in the Spring Term 2013
on the Recommendation of the Doctoral Committee

Prof. Dr. Lutz Jäncke
Prof. Dr. Volker Dellwo

Zurich, 2013

Contents

Abstract

Meanwhile, there is an increasing array of evidence, which shows that musicianship facilitates segmental, as well as supra-segmental aspects of speech processing. Based on this evidence, I present in this thesis two empirical studies investigating the influence of musical expertise on top-down and bottom-up processes of speech perception. The first study addresses stronger to top-down processes of speech perception, namely whether professional music training has an influence on the perceptual learning of speech sounds which were manipulated in terms of spectral- or envelope-related acoustic information. This study clearly showed a perceptual learning advantage due to musical expertise, but only for speech sounds containing their verbal information in the fine structure of the acoustic signal.

In the second study, I investigated the influence of musical expertise on bottom-up processes on speech perception. In a passive mismatch negativity (MMN) paradigm, I examined systematically differences in neural encoding of vowels and temporally manipulated CV syllables in a group of musicians and non-musicians. Results of this second study revealed that musicians elicited stronger brain responses on either temporarally (manipulation of the voice onset time, or the vowel duration) or spectrally (manipulation of pitch, or the second formant transition of the vowel) manipulated CV syllables.

The results of these two studies support the hypothesis that musical expertise advantages both, top-down processes as well as bottom-up processes of speech perception.

1 Introduction

In the early cognitive neurosciences, it was supposed that after ontogenesis motor- and sensory-related cortical and subcortical brain areas are incapable for alterations, whereas it was well accepted that brain regions like the hippocampus are more susceptible to functional and structural neural reorganisation processes (Jancke, 2009).

However, in the 1990th neuroscience first demonstrated that the auditory system of mammalian is highly susceptible to functional and structural changes (Jancke, 2009; Munte, Altenmuller, & Jancke, 2002; Schlaug, 2001). Cortical reorganization in the primary auditory areas could be shown in cats after cochlear lesions (Rajan, Irvine, Wise, & Heil, 1993) or in the monkeys' auditory cortex after a frequency discrimination training (Recanzone, Schreiner, & Merzenich, 1993).

In the field of human neuroscience, the effects of intense musical training on auditory processing have been well documented in various cross-sectional (Baumann, Meyer, & Jancke, 2008; Meyer et al., 2011a; Pantev, Engelien, Candia, & Elbert, 2001a; Schlaug, Jancke, Huang, & Steinmetz, 1995) and longitudinal studies (Hyde et al., 2009; Moreno et al., 2009). Professional musicians usually start practicing a musical instrument in very young years of their life by spending many hours per day playing their instrument. Though, professional musicians often show a lifetime training of many ten thousands of hours. This makes this group of individuals particularly fruitful for investigating neuroplastic alterations of the auditory system (Jancke, 2009; Munte, et al., 2002).

Anyhow, the investigation of professional musicians in contrast to non-musicians has shown an natural access for exploring the exceptional capability of the auditory system to underlie functional (Elmer, Meyer, & Jäncke, 2012; Oechslin, Van De Ville, Lazeyras, Hauert, & James, 2012; Ott, Langer, Oechslin, Meyer, & Jancke, 2011) and structural (Jancke, 2009; Munte, et al., 2002) changes (Hyde, et al., 2009; Moreno, et al., 2009).

In the meantime, various EEG studies found evidence for an increased sensitivity of the auditory system of professional musicians while they process timbre (Meyer,

Baumann, & Jancke, 2006), pitch (Bidelman, Krishnan, & Gandour, 2011), rhythm (Geiser, Sandmann, Jäncke, & Meyer, 2010), and duration (Marie, Kujala, & Besson, 2012) cues under attended or unattended experimental conditions. As an example, Pantev and colleagues (2001b) presented violin and trumpet tones to professional violinists and trumpet players. They revealed increased brain responses to the familiar and more intensively trained sound (i.e., timbre).

1.1 Musicianship and speech perception

Meanwhile, evidence showing that musicianship not only facilitates the processing of musical information, but that it influences also perceptual processes in domains that were not explicitly trained (Besson, Chobert , & Marie, 2011; Kraus & Chandrasekaran, 2010; Milovanov & Tervaniemi, 2011; Patel, 2011; Schön & Francois, 2011). Longitudinal studies provided evidence for a causal link between professional musical training and the competence of musicians to actively (Francois, Chobert, Besson, & Schon, 2012; Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Moreno et al., 2011; Moreno, et al., 2009) or passively (Chobert, Marie, Francois, Schon, & Besson, 2011) detect small spectral modulations in speech sounds, principally when acoustic cues are difficult to detect (Schön & Francois, 2011). Otherwise, a question that has not yet been intensively addressed is whether the cortical reorganizations previously observed in musicians may enable a functional advantage in encoding vowels. In fact, to the best of our knowledge this specific issue has previously only been investigated in 9-year-old children by Chobert and colleagues (Chobert, et al., 2011). Long-term experience in one domain influences acoustic processing in the other domain, though results can be interpreted as common acoustic processing.

Aniruddh Patel (2011) proposes a hypothesis to describe the conditions which must be met to drive adaptive plasticity from the musical domain to the domain of speech perception, called “The OPERA hypothesis”. Musical training benefits the neural encoding of speech. The OPERA hypothesis proposes that such benefits are driven by adaptive plasticity in speech-processing networks. That this plasticity occurs, Patel

proposes five necessary conditions: (1) **O**verlap: brain networks that processes acoustic feature used in speech and music must overlap, (2) **P**recision: music places higher demands on these shared networks than does speech, in terms of the precision of processing, (3) **E**motion: the musical activities that engage this network elicit strong emotion, 4) **R**epetition: the musical activities that engage this network are frequently repeated, and (5) **A**ttention: the musical activities that engage this network are associated with focused attention. It is important to mention, that these fife conditions of the OPERA hypothesis are not automatically met, but that professional musicians mostly better meet these conditions than amateur musicians. In fact, even though speech and music are characterized by different psychoacoustic features, both signals convey information by means of timing, pitch, and timbre cues (Kraus and Chandrasekaran, 2010). There are more acoustic similarities between these two domains than differences and a growing number of neuroscientists are interested in investigating transfer effects from musical training to the processing of speech sounds. Most of the studies investigating transfer effects of musical expertise on speech perception address the question whether musical expertise may favour segmental (Elmer, Meyer, & Jancke, 2011; Marie, Magne, & Besson, 2011; Ott, et al., 2011) or supra-segmental (Marie, et al., 2011; Marques, Moreno, Castro, & Besson, 2007; Oechslin, Meyer, & Jancke, 2010; Schon, Magne, & Besson, 2004) aspects of elemental speech processing. Moreover, neural reorganisation processes have previously not only been shown to be limited to the auditory cortex, but more likely to engage huge parts of cortical (Besson, et al., 2011; Milovanov & Tervaniemi, 2011; Patel, 2011) and subcortical (Kraus & Chandrasekaran, 2010; Kraus, Skoe, & Parbery-Clark, 2008) tissue. Exactly these large-scale neuronal reorganization processes, in association with the partial overlap between speech and music functions (Patel, 2011), motivated a growing number of scientists to investigate relationships between neuroplasticity and speech processing in professional musicians (Jäncke, 2012). This kind of research provided evidence for a causal link between professional musical training and the competence of musicians to actively (Marie, et al., 2011) or passively (Boh, Herholz, Lappe, & Pantev, 2011; Marie, et al., 2012; Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005) detect small spectral modulations in speech sounds, principally when acoustic cues are difficult to detect (Schön & Francois, 2011).

Meanwhile, there is evidence from cross-sectional studies showing that professional musicians are advantaged over non-musicians in detecting temporal modulations in CV syllables or even single words, as provided by previous fMRI (Elmer, et al., 2012) and EEG (Besson, Schon, Moreno, Santos, & Magne, 2007; Kühnis, Elmer, Meyer, & Jancke, 2012; Marie, et al., 2011; Ott, et al., 2011) studies.

To date, there is also growing evidence indicating that musicianship not only favours perceptual processes, but that it is indeed associated with behavioural advantages in several cognitive domains (Baumann, et al., 2008).

1.2 Transfer effects of musical training

A variety of studies demonstrated that musical expertise not only promote function in the trained field but also influences other domains such as cognition (Aleman, Nieuwenstein, Bocker, & de Haan, 2000; Bilhartz, Bruhn, & Olson, 1999; Brochard, Dufour, & Despres, 2004; Ho, Cheung, & Chan, 2003; Sluming, Brooks, Howard, Downes, & Roberts, 2007). A research question that is of fundamental interest in association with musical training and transfer effects is whether musicianship exerts an influence on perceptual learning mechanisms. Some studies investigated the influence of musical expertise on verbal learning (Ho, et al., 2003; Schlaug, Norton, Overy, & Winner, 2005). Nevertheless, it remains unclear exactly which of the trained acoustic parameters (i.e., frequency, amplitude, timing, or timbre) are relevant for facilitating transfer effects from musical training to cognitive aspects of speech processing. Some previous work illustrating a more efficient processing of spectral information in musicians (Besson, et al., 2007; Boh, et al., 2011; Draganova, Wollbrink, Schulz, Okamoto, & Pantev, 2009; Gaab et al., 2005; Schon, et al., 2004), others found efficient processing of temporal information in musicians (Besson, et al., 2011; Marie, et al., 2011).

Anyhow, these studies demonstrate that musical expertise benefits perceptual learning in speech. It is still an unanswered question whether musicians possess a behavioural advantage over non-musicians in learning auditory-presented verbal material that is characterized by spectral information only or whether musicians

show behavioural advantages when speech-stimuli are characterized by envelope-related acoustic information only.

However, musical expertise would not only advance top-down processes as posed before but also bottom-up processes. To investigate bottom-up processes passive auditory paradigms are often considered to be the better indicators to objectify in audition (Naatanen & Alho, 1995). There is growing evidence from a variety of EEG studies investigating modulation of bottom-up processes in the domain of speech in professional musicians by reporting enhanced electrophysiological responses on voice onset modulations in consonant-vowel (CV) syllables in a passive paradigm (Chobert, et al., 2011). Further, Nikjeh and colleagues (2009) reported in a passive listening task earlier electrophysiological responses to CV syllables which differ in place of articulation.

1.3 Division of labour between the auditory cortices

In the second half of the 19th century, the investigation of lesion patients brought first evidence for the localisation of the speech processing areas. After the death of persons with speech impairments, their brains were examined and left hemispheric lesions were connected to the deficits in language and speech processing. In this time, Broca and Wernicke conducted great work. Paul Broca (1861) related severe speech production deficits with a lesion in the left inferior frontal gyrus, now known as Broca's area. Shortly after Broca, Carl Wernicke (1874) related damage of the left posterior superior temporal gyrus, now known as Wernicke's area, to lexical impairments. For more than one hundred years, it was general knowledge, that speech function, namely speech production and speech perception were strongly located in the left hemisphere in the Wernicke and the Broca area. But in 1997, Binder and colleagues published an fMRI study showing clearly regions outside the classical language areas to be involved in language processes. This work marked a beginning of a new discussion about the localisation of speech functions. Over time, evidence accumulated showing that the left and the right planum temporale both are strongly involved in processing sounds and speech (Gandour et al., 2003; Jancke,

Wustenberg, Scheich, & Heinze, 2002a; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004). Moreover, a meta-analysis by Vigneau and colleagues (2011) pronounced the involvement of the right hemisphere in a variety of language-related processes.

A milestone in the field of speech processing was probably the proposition of a division of labour between the auditory cortices of the two hemispheres (Poeppel, 2003; Zatorre & Belin, 2001; Zatorre, Evans, Meyer, & Gjedde, 1992). In this context, the innovative idea was that the left auditory cortex is more likely involved in processing fast changing and transient acoustic cues, whereas its right counterpart is preferentially driven by slow acoustic modulations. This framework implies, for example, that the voice-onset time (VOT) which can be defined as a fast changing acoustic cue represented by the time between the release of a stop consonant and the onset of vocal folds vibrations (Lisker & Abramson, 1964), more strongly relies on left-sided auditory cortex functions (McGettigan & Scott, 2012; Zaehle, Jancke, Herrmann, & Meyer, 2009). Otherwise, spectral manipulations of the acoustic signal (i.e., long integration windows, fundamental frequency and formant transitions of vowels) more likely engage right-sided auditory cortex functions (Abrams, Nicol, Zecker, & Kraus, 2008; Johnsrude, Penhune, & Zatorre, 2000; Meyer, 2008). This framework has been supplemented by previous electrophysiological work describing slow and fast spontaneous oscillations originating from auditory-related cortical fields during resting state (Giraud et al., 2007).

2 Methods

For all my studies, I used one of the oldest techniques to measure electrical brain activity, namely the electro encephalogram (EEG). This technique was first developed by Hans Berger in the year 1929 and is today, thanks to improved mathematical implementations, a well-established technique in the field of neuroscience. The EEG gained popularity, as it was, until the introduction of PET and fMRI, the only method in cognitive neuroscience to noninvasively study brain activity

For measuring electrical brain activity, 32 up to 256 electrodes are placed on the skin of the scalp. Voltage fluctuations between each electrode and a reference are recorded. These voltage fluctuations reflect synchronous electrical activations originating from large samples of nerve cells within the brain. The EEG technique has some giant advantages over other techniques such as MRI.

First, EEG is characterized by an excellent temporal resolution in the time range of milliseconds. This opens the possibility to have an important view inside the dynamics of brain processes. Second, EEG measurement does not emit any noise, though it is predestined for auditory investigations.

Due to novel mathematical approaches, the analysis of EEG opens a wide range to look deeper in the dynamics for brain processes (microstate analysis) or to localize activations in the brain with low resolution tomography (LORETA).

2.1 Event related potentials

Event related potentials (ERPs) can be measured by repeatedly presenting a stimulus to a person and then averaging the time locked signal. With this procedure, electrical brain activity that is evoked by the stimulus can be separated from other, unrelated signals which were statistically eliminated by averaging. In contrast to spontaneous EEG data, ERPs are characterized by rather small amplitudes in the range of some micro volts (μV). To get an acceptable signal to noise ratio, a stimulus must be presented to a volunteer up to 80 times. Auditory evoked potentials (AEP) are in this context of special interest. AEPs were evoked by presenting auditory stimuli to a

participant. AEPs are characterized by different positive and negative deflections. A large variety of components are labelled by either their latency and polarity (i.e. N100), their localization on the scalp (e.g. early left anterior negativity) or their functional significance (e.g. mismatch negativity).

Most prominent components in AEP are the N1- and the P2 component. On the one hand, the N1 reaches an amplitude maximum in the negative voltage range at about 100ms after stimulus onset. The N1 is related to basic auditory processing (Meyer, et al., 2006; Naatanen & Picton, 1987) and is influenced by basic stimulus properties (i.e. Mulert et al., 2005). Moreover, the N1 is influenced by auditory experience, for instance professional musical training (Baumann, et al., 2008; Pantev, et al., 2001a). Meanwhile, it is established that N1- as well as the P2-component of the AEP, reflect electrical activity originating from auditory-related brain regions (Vaughan & Ritter, 1970).

On the other hand, the P2 component reaches its positive voltage maximum in a time range of about 200ms after stimulus onset. Today, the meaning of the P2 component is not fully clear. There is some evidence indicating that top-down processes are often associated with the modulation of the auditory-evoked P2 component (Hillyard, 1981; Picton & Hillyard, 1974; Schwent & Hillyard, 1975).

Further, the mismatch negativity (MMN) represents an ERP component which is elicited by an unexpected stimulus (deviant) in a series of expected stimuli (standard). The MMN was first discovered by Risto Näätänen (for an overview see Näätänen, 1992). In a classical MMN paradigm, a frequent standard stimulus and a seldom deviant stimulus were presented to a volunteer. To both, the standard and the deviant stimulus, the volunteer elicits ERPs, where the ERP to the deviant stimulus is characterized by a stronger negativation 100ms after aberration between the two stimuli. This negativation is called mismatch negativity and maintains for 200-250ms. Interestingly, the magnitude of the MMN correlates with the magnitude of the aberration between deviant and standard stimuli. Furthermore, the MMN is independent of attention. Though, the MMN represents a preattentive ERP component, stronger influenced by bottom-up then by top-down processes.

AEPs often build the basis for further analysis such as the microstate analysis and the estimation of activation in the brain, LORETA.

2.2 Microstate analysis

With component analysis, only one electrode or a pool of electrodes are investigated and analysed. This analysis is blind for changes in electrodes which are not included in the analysis.

The morphology of EEG waveform strongly depends on the position of the recording electrode. One major problem is the fact that one electrode shows a classical component other electrodes might show a clear component at a different latency or even another polarity. For example, in an AEP, the N1 component results in a minimum voltage at the Cz position on the top of the head at 100ms after stimulus onset, but at the mastoid position (TP9 or TP10) at the same latency the scalp voltage results in a maximum. This finally means that the analysis of the EEG data also depends on the selection of the analysed electrodes. To solve this problem, it is necessary to have a closer and more holistic view on the EEG data.

Scalp potential distributions or EEG maps are one possibility to holistically describe EEG data. For ERP components, each electrode codes the scalp voltage for a time point and a position on the scalp. Scalp voltages between the electrodes will be interpolated. This approach opens up the possibility to illustrate EEG data topographically. Further, Dietrich Lehmann (1987) showed that series of EEG maps are typically characterized by sequential periods of stable map topographies separated by short transitions. Periods of stable map topographies are called microstates. In an ERP or an AEP, microstates represent a certain information processing step that leads from perception to action (Michel, 2009a). The microstate segmentation of the AEP proposes to define components in terms of the sequentially appearing map topographies instead of the sequentially appearing peaks at certain electrodes. This means that waveform morphology is not a defining attribute of a component any more, only the spatial configuration of the scalp voltage distribution.

A spatial cluster analysis can be used to define the most dominant topographies in a given AEP map series. Further, the fitting by means of spatial correlation allows to define when each of these maps is presented in the data (Murray, Brunet, & Michel, 2008; Pascual-Marqui, Michel, & Lehmann, 1995).

Finally, the influence of experimental variables on the appearance, the temporal sequence, the duration, and the strength of the microstates can be tested statistically (Michel, Seeck, & Landis, 1999). This approach has been introduced as topographic component recognition (TCR) by Brandeis et al. (1992).

Basing on the hypothesis that each microstate represents a certain information processing step (Michel, 2009a), it would be obvious to estimate the putative sources of each of the different map topographies. This can be done by using a method developed by Pascal-Marqui et al. (2002), namely the Low Resolution Brain Electomagnetic Topography (LORETA).

2.3 LORETA

LORETA (Pascual-Marqui, et al., 2002) is an electrical brain imaging technique to estimate putative brain sources. LORETA computes 3D linear estimations for the EEG inverse problem. It was previously shown that LORETA permits to reliably estimate the current density maxima of AEP components (Meyer, et al., 2006; Pascual-Marqui, et al., 2002; Pascual-Marqui, Michel, & Lehmann, 1994). This approach, unlike conventional dipole fitting, does not require a-priori assumptions about the number and the localization of the dipoles. LORETA calculates the three dimensional distribution of electrically active neuronal generators in the brain as a current density value (A/m^2), and provides a solution for the inverse problem by assuming that the smoothest of all possible activity distributions is the most plausible one for explaining the data. The characteristic feature of this particular inverse solution approach is the low spatial resolution, which conserves the location of maximal activity, but with a certain degree of dispersion (Mulert et al., 2004). LORETA determines the current density distribution for epochs of brain electrical activity on a dense grid of 2394 voxels at 7 mm spatial resolution. The localisation error of LORETA's source

identification may vary between 7 (Pascual-Marqui, et al., 1994) and 14 mm (Phillips, Rugg, & Friston, 2002). LORETA refers to a three-shell spherical model registered to the Talairach human brain atlas. The source estimations are provided as x, y, z coordinates and are situated relative to the intercommissural line (AC–PC line) in horizontal (x), the anterior/posterior (y), and vertical (z) directions. The solution space is confined to the gray matter portion of the human cortex, which rules out the option that subcortical tissue and white matter contribute to the solution.

3 Aims of this thesis

This thesis is located in the scientific field of functional neural plasticity of speech perception. In two studies, I investigated the perception of speech stimuli in professional musicians and non-musicians.

First question:

Do musicians show a perceptual learning advantage in contrast to non-musicians, when they hear words limited in spectral or temporal information only?

Study one aimed to examine perceptual learning differences depending on musical expertise. In this study, I presented pseudo words. This pseudo words were manipulated in that way, that they contained their verbal information either in their temporal envelope or in their spectral fine structure. Interestingly, musicians outperformed non-musicians clearly, but only when they heard the pseudo word containing their verbal information in the fine structure. Musicians did not show any advantage in perceptual learning over non-musicians when they heard pseudo words containing their verbal information just only in their temporal envelope.

Second question:

Do musicians have an advantage to better recognize differences in manipulated CV syllables?

The second study aimed to examine perceptual advantages of professional musicians in a passive listening paradigm. I produced semi-artificial CV syllables manipulated in vowel duration, pitch and in the second formant transition, as well as in the voice onset time. Most interestingly, musicians elicited stronger electrophysiological responses to all CV manipulations.

Third question:

Is it possible in the context of the AST (asymmetric sampling in time hypothesis by Poeppel, 2003) to show any lateralisation effect in processing CV syllables manipulated in temporal and spectral features?

The **second study** revealed as predicted by the AST, a lateralization in processing CV features. A LORETA analysis revealed stronger activation of the left auditory cortex when processing differences in vowel duration and VOT manipulation. Further, the processing of manipulation in the second formant of the vowel and of vowel pitch activated stronger brain regions of the right auditory cortex.

Fourth question:

Is it possible from a methodological point of view to show any perceptual learning effect in a microstate analysis?

The first study aimed to demonstrate the advantage of microstate analysis in a classic paradigm investigating musicians and non-musicians. Most interestingly, the microstate analysis brought prolonged P2-like microstate in musicians to light.

4 Empirical studies

4.1 First Study – Perceptual learning and musical expertise

Title of the original study:

Musicianship Boosts Perceptual Learning of Pseudoword-Chimeras: an Electrophysiological Approach.

Jürg Kühnis, Stefan Elmer, Martin Meyer, Lutz Jäncke

Published in Brain Topography, 26(1), 110-25

4.1.1 Abstract

A vast amount of previous work has consistently revealed that professional music training is associated with functional and structural alterations of auditory-related brain regions. Meanwhile, there is also an increasing array of evidence, which shows that musicianship facilitates segmental, as well as supra-segmental aspects of speech processing. Based on this evidence, we addressed a novel research question, namely, whether professional music training has an influence on the perceptual learning of speech sounds. In the context of an EEG experiment, we presented auditory pseudoword-chimeras, manipulated in terms of spectral- or envelope-related acoustic information, to a group of professional musicians and non-musicians. During EEG measurements, participants were requested to assign the auditory-presented pseudoword-chimeras to one out of four visually presented templates. As expected, both groups showed behavioural learning effects during the time course of the experiment. These learning effects were associated with an increase in accuracy, a decrease in reaction time, as well as a decrease in the P2-like microstate duration in both groups. The musicians showed an increased learning performance compared to

the controls during the first two runs that pertained to the spectral condition. This perceptual learning effect, which varies as a function of musical expertise, was reflected by a reduction of the P2-like microstate duration. Results may mirror transfer effects from musical training to the processing of spectral information in speech sounds. Hence, this study provides evidence for a relationship between changes in microstates, musical expertise, and perceptual learning mechanisms.

4.1.2 Introduction

In the last two decades, neuroscientific evidence has demonstrated that the mammalian auditory system is highly susceptible to functional and structural changes (Jancke, 2009; Munte, et al., 2002; Schlaug, 2001). For example, several animal studies provide evidence for cortical reorganization in primary auditory areas of adult cats after cochlear lesions (Rajan, et al., 1993) or in the monkeys' auditory cortex after a frequency discrimination training (Recanzone, et al., 1993). In the field of human neuroscience, the effects of intense musical training on auditory processing have been well documented in various cross-sectional (Baumann, et al., 2008; Meyer, et al., 2011a; Pantev, et al., 2001a; Schlaug, et al., 1995) and longitudinal studies (Hyde, et al., 2009; Moreno, et al., 2009). Since professional musicians usually start practising a musical instrument at an early stage of their life and usually spend many hours per day playing their musical instrument, this specific group of individuals are particularly fruitful for investigating neuroplastic alterations of the auditory system (Jancke, 2009; Munte, et al., 2002).

Meanwhile, evidence shows that musical training not only facilitates the processing of musical information, but that it also effects perceptual processing in domains that have not been explicitly trained (Besson, et al., 2011; Kraus & Chandrasekaran, 2010; Milovanov & Tervaniemi, 2011; Patel, 2011; Schön & Francois, 2011). In fact, even though speech and music are characterized by different psychoacoustic features, both signals convey information by means of timing, pitch, and timbre cues (Kraus and Chandrasekaran, 2010). There are more acoustic similarities between these two domains than differences and a growing number of neuroscientists are interested in

investigating transfer effects from musical training to the processing of speech sounds. Most of the studies investigating such transfer effects address the question whether musical expertise may favour segmental (Elmer, et al., 2011; Marie, et al., 2011; Ott, et al., 2011) or supra-segmental (Marie, et al., 2011; Marques, et al., 2007; Oechslin, et al., 2010; Schon, et al., 2004) aspects of elemental speech processing.

To date, there is also growing evidence indicating that musicianship not only favours perceptual processes, but that it is indeed associated with behavioural advantages in several cognitive domains (Baumann, et al., 2008). Such benefits are apparent even in individuals who have not been explicitly trained (Aleman, et al., 2000; Bilhartz, et al., 1999; Brochard, et al., 2004; Ho, et al., 2003; Sluming, et al., 2007). A research question that is of fundamental interest in association with musical training and transfer effects is whether musicianship exerts an influence on perceptual learning mechanisms. Even though a comparable research question has been previously addressed by other authors (Ho, et al., 2003; Schlaug, et al., 2005), it remains unclear exactly which of the trained acoustic parameters (i.e., frequency, amplitude, timing, or timbre) are relevant for facilitating transfer effects from musical training to cognitive aspects of speech processing. Based on previous work illustrating a more efficient processing of spectral information in musicians (Besson, et al., 2007; Boh, et al., 2011; Draganova, et al., 2009; Gaab, et al., 2005; Schon, et al., 2004), one would expect musicians to possess a behavioural advantage over non-musicians when it comes to learning auditory-presented verbal material that is characterized by spectral information only. To date, it is unknown whether musicians show behavioural advantages when speech-stimuli are characterized by envelope-related acoustic information only.

Even though the electrophysiological (EEG) technique has a long tradition in the field of cognitive neuroscience, recent novel applications and mathematical implementations have enabled researchers to address more complex questions pertaining to how the human brain functions. As regards the auditory system, most of the EEG studies focussed on the auditory-related N1-P2 complex or the associated latency bands which are known to reflect the auditory representation and energy of speech and non-speech sounds (Meyer, et al., 2006; Ostroff, Martin, & Boothroyd,

1998; Pratt, Starr, Michalewski, Bleich, & Mittelman, 2007; Sharma, Marsh, & Dorman, 2000). For example, previous EEG studies demonstrate an increased auditory representation of musical sounds in musicians compared to non-musicians (Baumann, et al., 2008; Besson, et al., 2007; Meyer, et al., 2011a; Schlaug, et al., 2005). Results pointing to increased cortical representation of the primary instrument played by musicians have also come to light (Meyer, et al., 2011a; Pantev, et al., 2001b; Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005a).

A particularly interesting approach known as Topographic Pattern Analysis was recently proposed by Murray and colleagues (2008). This methodological procedure is helpful in that it permits to evaluate and interpret EEG data by performing a step-by-step analysis. Even though the general principles of the microstate segmentation have been described extensively elsewhere (for a comprehensive overview consider Murray et al., 2008; Michel et al., 2009), in turn we will provide a short summary of this procedure. The Topographic Pattern Analysis can be used for evaluating complex topographic information arising from high-density EEG measurements. In particular, it permits to determine electrophysiological differences between experimental conditions and groups with regard to field strength, topography, and latency shift. This specific method implies that evoked activities are more or less stable scalp voltage topography manifestations lasting for several tens of milliseconds before a qualitative change in the configuration of the electric fields occur. Such a qualitative change leads to another topographic map, which remains stable for a certain duration. These stable topographic maps are called microstates and can be determined by using a topographic atomize and agglomerate hierarchical cluster analysis (T-AAHC) on the grand-averaged ERP's across conditions and groups.

In the present work we performed an EEG study specifically designed to evaluate perceptual learning mechanisms as a function of musical expertise. With this purpose in mind we presented pseudoword-chimeras characterized by spectral- and envelope-reduced information to musically trained and untrained individuals. In accordance with previous studies showing that musicians are principally facilitated in processing spectral information (Besson, et al., 2007; Boh, et al., 2011; Draganova, et al., 2009; Gaab, et al., 2005; Schon, et al., 2004), we expected to obtain the following

results: the musicians will generally show an increased learning performance at recognizing the auditory chimeras that are only defined by spectral information. In addition, we expect the musicians to exhibit a faster learning curve during the time course of the experiment. Additionally, due to our consideration of previous work indicating that musicianship favours the processing of several different psychoacoustic dimensions (Kraus & Chandrasekaran, 2010), such as, frequency- (Meyer, et al., 2011a), duration-, as well as amplitude-modulations (Lee, Skoe, Kraus, & Ashley, 2009), we expect to find that musicians are even better at recognizing pseudoword-chimeras that are only characterized by amplitude information.

With regard to the electrophysiological data, we expect to find increased microstate-related global field power (GFP) magnitudes of the N1-P2 complex in musicians in response to both spectral- and envelope-reduced pseudoword-chimeras. Furthermore, we hypothesize that the better learning performance of the musicians shall be reflected by an alteration of the N1-P2-like microstates duration. The reason for focussing on the N1-P2 complex is supported by the fact that these auditory-evoked responses have been shown to constitute an objective method for measuring the processing of any discrete acoustic feature (Hyde, 1997). In fact, previous work has consistently shown that acoustic expertise in musicians is associated with increased cortical representations of speech (Marie, et al., 2011) and non-speech sounds (Pantev et al., 1998), as reflected by enhanced brain responses. Consequently, one may expect that changes in any discrete acoustic features of the stimuli, as well as expertise, will be associated with a modulation of the microstates duration during the processing stage of the N1-P2 complex. This expectation is supported by earlier work indicating a relationship between microstates duration and brain processes (Lehmann & Michel, 2011).

4.1.3 Material and methods

4.1.3.1 Participants

Fifty volunteers with normal audiological status and no history of neurological diseases participated in this study. All participants were native Swiss-German or

German speakers and were strongly right-handed according to standard handedness questionnaires (Annett, 1970; Jancke, 1996). The first group of participants consisted of twenty-five professional musicians (13 men and 12 women) in the age range of 19 to 47 years (mean age 25.3 ± 5.9 SD). Musicians had more than ten years of musical training. All musicians commenced their musical training before the age of seven years old (mean age 6.5 year ±1.3 SD) and practised their musical instrument for 3.2 hours per day on average (±1.5 SD). The musician group consisted of seven singers, five pianists, and four participants who play classical drums. The control group consisted of twenty-five volunteers without formal musical education in the age range of 19 to 45 years (mean age 26.4 ± 7.5 SD). All participants gave informed written consent in accordance with procedures approved by the local ethics committee and were paid for their participation.

4.1.3.2 Behavioural data

4.1.3.2.1 Musical aptitudes

In order to measure the musical aptitudes of each participant, we applied the "Advanced Measure of Music Audition" (AMMA) test implemented by Edwin E. Gordon (1989). This specific test is based on the assumption that holding auditory-presented musical patterns in memory and detecting both melodic and rhythmic variations constitutes a fundamental prerequisite for musical aptitude. During the AMMA test volunteers heard short pairs of piano sequences (both pieces have a duration of about five seconds) and were required to decide whether these sequences were equivalent, rhythmically different, or tonally different.

4.1.3.2.2 Cognitive capability

In order to exclude that possibility that putative differences in intelligence between the two groups could impact the data in some direction, we adopted two short intelligence tests: the KAI (Lehrl, Gallwitz, Blaha, & Fischer, 1991) and the MWT-B

(Lehrl, 1977). Table 1 gives an overview of the biographical and behavioural data of the participants.

Table 1: Biographical and behavioural data of the two groups (M = musician, NM = non-musicians, AMMA = Advanced Measures of Music Aptitude, KAI = Kurztest der aktuellen geistigen Leistungsfähigkeit, MWT = Mehrfachwahl-Wortschatz-Intelligenz). ** Significant difference between the two groups, p = .004 ($F_{(1, 48)} = 8.984$).

	Age [a]		Gender		KAI [IQ]		MWT-B [IQ]		AMMA** [PR]	
	m	sd	m	f	m	sd	m	sd	m	sd
M	25.28	5.669	12	13	127.76	10.82	113.10	13.58	69.16	14.72
NM	26.44	7.556	12	13	123.74	12.22	115.48	14.21	55.76	16.82

4.1.4 Stimuli

The auditory stimuli we presented to the participants consisted of forty phonotactically legal low associative pseudowords taken from the verbal / non-verbal (VLT / NVLT) learning test (Sturm & Willmes, 1999). All stimuli used in this test are complex disyllabic pseudowords (example: “hentes”, “fasenz”, “trelas”). The stimuli were spoken by a professional male speaker and recorded at a sampling rate of 44.1 kHz. All pseudowords matched in intensity (amplitude normalization with the PRAAT software, http://www.fon.hum.uva.nl/praat/); the duration of the auditory stimuli ranged from 900 to 1000 ms.

We produced four different speech-noise chimeras for each of the auditory stimuli by using the MATLAB software (versionR2010a, http://www.mathworks.com). The first step was to create noise analogues of the original stimuli by maintaining the spectral density and intensity. The second step entailed filtering both sound signals (i.e., the original stimuli and the noise analogues we created) with two, three, or eight complementary band-pass filters in the frequency range from 80 Hz to 17.64 kHz (0.4 x sampling frequency). We did this by using nearly rectangular frequency filters with logarithmic frequency spacing (Shen, 2001; Smith, Delgutte, & Oxenham, 2002). The

purpose of this procedure was to diversify the intelligibility of the auditory stimuli by manipulating the number of band-pass filters (Smith, et al., 2002). In the third processing step, the filtered signals from the corresponding frequency bands of the two signals (i.e., speech- and noise-stimuli) were converted by using a chimerizer procedure that interchanges the envelope (i.e., the time course of the signal amplitude) and the fine structure (i.e., the time course of the spectral information) of the two input signals by producing two partial speech-noise chimeras. We used the Hilbert Transform to factorize envelope and fine structure from the two input signals (speech- and noise-stimuli). With this purpose in mind, we modulated the fine structure and the envelope of the speech and noise signals (and vice versa), in order to obtain partial speech-noise chimeras (two partials chimeras per frequency band in total). Finally, we summed up these partials chimeras over all frequency bands, which resulted in two multi-band chimeras for each of the three band-pass filtered input signals (i.e., two, three, or eight band-pass filters). Whereas the first chimera was characterized by the envelope of the speech sound and by the fine structure of the noise signal, the second one was composed of the envelope of the noise sound, a straight line, and the fine structure of the speech signal (Shen, 2001). A sample of these auditory stimuli can be downloaded under: (http://www.neurowissenschaft.ch/mmeyer/.

The visually presented words (one target and three distractors) consisted of pseudowords, which were phonological neighbours of the target word. These phonological neighbours were collected in the context of a pre-test study conducted with four students who were instructed to write down the pseudowords they heard. In this context, we used the erroneously reported pseudowords as distractors for the main experiment. By applying this procedure, we only focused on distractors that exhibited a similar phonological structure (example: target word: “hentes”; distractors: “hintis”, “temtis”, “sentis”).

Before running the EEG experiment, we performed a short behavioural experiment with 17 volunteers, in order to estimate the intelligibility of the speech-noise chimeras. Whereas all of the auditory pseudoword chimeras were presented via headphones, the target words and the distractors were visually presented on a

computer screen. In this context, volunteers were instructed to find the pseudoword chimera that was presented auditorily and to select the target by pressing one of four response buttons. This procedure was the same as the method we employed in the EEG experiment. One difference is that we recruited two different groups of individuals for performing the pilot study and the EEG experiment.

This behavioural approach revealed an intelligibility (i.e., number of correctly recognized items) of about 37% for the chimeras containing speech information in the envelope and filtered with three band filters (ENV3), as well as for the chimera containing speech information in the fine structure and filtered with three band filters (FS3). The other two chimeras, namely, the chimera that was defined by the speech information in the envelope and filtered with two band filters (ENV2), as well as the chimera containing speech information in the fine structure and filtered with eight band filters (FS8) exhibited an intelligibility of about 30%. As expected, the original and not the manipulated pseudowords were characterized by an intelligibility of nearly 100%.

4.1.5 Experimental procedure and task

During the EEG measurement, participants sat in a comfortable chair at a distance of about 75 cm from a 19-inch-monitor and were instructed to look at a small fixation cross that was presented on the screen. This was instructed in order to avoid movement-related artefacts during data acquisition. Every participant performed four runs, each of which lasted about 15 minutes. During each run, the stimuli of all five conditions (ENV2, FS8, ENV3, FS3, and original pseudowords that acted as a control condition) were presented 150 times in a randomized order at a sound pressure level of about 75 dB (Digital Sound Level Meter 329, Voltcraft) with Sennheiser in-ear HIFI-headphones (CX-350). Two seconds after stimulus onset, the pseudowords (target word) and three other pseudowords, which acted as distractors were visually presented on the screen. Participants were instructed to recognize and select the pseudoword (target word) they heard by pressing one of four possible

response buttons with the middle- and forefingers of both hands (randomized across the participants).

After response selection, a delay of one second followed before the next auditory stimulus was presented. The presentation of the auditory stimuli and the recording of behavioural responses were controlled by the Presentation software (www.neurobs.com; version 14.5). We collapsed the behavioural and electrophysiological responses of the FS3 and FS8 into one FS condition for all of the analysis in this work. This was done because of two reasons, namely, we were interested in analysing the influence of expertise on the processing of spectral- and enveloped-reduced pseudowords and we also wanted to increase the variance across the two groups. The same procedure was applied for the ENV3 and ENV2 stimuli (ENV condition).

4.1.6 EEG acquisition and pre-processing

A continuous EEG was recorded by the high-density Geodesic EEG system (Electrical Geodesics, Inc., USA) using 256 scalp electrodes (impedance < 40kΩ, online vertex reference; sampling rate of 500 Hz; band-pass filtered at 0.1-250 Hz). Brain Vision Analyser software (Version 2.01, Brainproducts, Munich, Germany) was used for all pre-processing steps. We reduced the number of electrodes from 256 to 204 by removing the electrodes situated on the cheeks and on the neck. Data were filtered off-line from 0.1 to 150 Hz. Artefacts (i.e., eye movements and eye blinking) were eliminated by using an independent component analysis (ICA) (Jung et al., 2000) in association with a semi-automatic raw data inspection. Furthermore, noisy data collected from distinct electrodes were interpolated (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987), that is, data were re-referenced off-line to an average reference, and sectioned into 500 ms segments (from 100 ms pre-stimulus to 400 ms post-stimulus). A baseline correction relative to the -100 to 0 ms pre-stimulus time period was applied and all segments were averaged for each participant, each condition, and each run, in order to compute auditory evoked potentials (AEPs). In addition, multi-subject grand averages were calculated for each group and condition.

4.1.7 Data Analysis

4.1.7.1 Behavioural data

First, we checked whether participants correctly performed the recognition task by averaging the responses to the clean, presented pseudowords for each participant. All participants attained recognition performances between 88% and 100% (mean 96.5%).

Following the lead of previous studies (Luo & Poeppel, 2007; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Smith, et al., 2002), we collapsed the number of correct ENV2 and ENV3 responses (accuracy ENV condition), as well as the FS8 and FS3 responses (accuracy FS condition) for each participant and run. We analysed the data in this way, in order to estimate general group differences while processing "envelope" (ENV) and "fine structure" (FS) stimuli. The same procedure was applied for evaluating reaction time (RT).

The evaluation of RT was performed by using a mixed-effect regression model analysis. Otherwise, since accuracy data were binomially distributed, we performed a generalized linear mixed model for binomially distributed outcome. For both analyses, we used *run* (run 1, 2, 3, 4), *condition* (ENV, FS), and *group* (musician, non-musician) as fixed factors, as well as *participants* and *items* as random factors (Baayen, Davidson, & Bates, 2008; Clark, 1973; Locker, Hoffman, & Bovaird, 2007). The main problem when using standard ANOVAs and regression models is that these analyses are conducted as either by-subjects or as by-items; however, the mixed-effects model allows subject and item effects to be included in the same model. This in turn potentially solves the language-as-fixed-effect fallacy (See Brysbaert (2007) for a primer on mixed-effect models). The model was then fitted using the restricted maximum likelihood estimation (REML).

4.1.7.2 EEG data

In the present work ERP data were analysed by using a multistep procedure which bases on the analysis of response topography and response strength which is useful

to differentiate between underlying generators as a function of expertise and task. Furthermore, this procedure enables to disentangle latency shifts of brain activity across experimental conditions and groups from modulation in the strength of responses of statistically indistinguishable brain generators (Michel, 2009b; Murray, et al., 2008). This kind of analysis is independent of the reference electrode (see, e.g. Michel et al., 2004) and is insensitive to amplitude modulations. All EEG analyses were performed by using the Cartool software package (Version 3.43/869; The Cartool Community group, sites.google.com/site/cartoolcommunity).

4.1.7.2.1 Global field power analyses (GFP)

By conducting the first step of the stepwise evaluation, which involves defining the time windows of interest, we compared GFP differences between the two groups by collapsing both conditions (ENV, FS) for each participant together into one auditory-evoked-potential (AEP) and by computing GFP time courses. In order to control for inter-individual variance across participants, we normalized the data to the mean GFP for each participant. For each electrode we marked the first time point when the previously calculated t-test between the two groups exceeded the statistical alpha criterion of .05 for at least ten consecutive data points (i.e., time windows which were larger than 20 ms). This criterion is an adequate measure for describing AEP modulations (Fort, Delpuech, Pernier, & Giard, 2002; Murray et al., 2004; Murray et al., 2002).

4.1.7.2.2 Microstate analyses

In the second step, we performed microstate analyses. The logic behind this analysis was to identify spatio-temporal differences in scalp voltage distributions between the groups and the two conditions (i.e., FS and ENV). With this purpose in mind, we applied a “topographic atomize and agglomerate hierarchical cluster” analysis (T-AAHC) on the grand-averaged AEP’s for each condition and group over a segment length of 400 ms. The T-AAHC analysis is used for the temporal segmentation of the EEG data. It functions by identifying periods of stable electric field topographies or

"functional microstates". This segmentation procedure generates hypotheses about when topographic modulations actually occur. The periods of stable topography are functionally relevant periods of interest (https://sites.google.com/site/cartoolcommunity/). This procedure incorporates Krzanowski-Lai criterion to identify the optimal number of template maps that can explain the data (Krzanowski & Lai, 1988; Murray, et al., 2008; Tibshirani & Walther, 2005).

The time window we used for statistical analysis primarily bases on the hypothesis that the processing of the perceptual properties of acoustic stimuli as a function of expertise is most likely reflected by a modulation of the N1-P2 complex. In order to validate this working hypothesis we additionally performed GFP comparisons between the two groups. Since the GFP analysis confirmed a modulation of brain responses overlapping with the auditory-evoked N1-P2 complex, for the microstates analyses we focused on the latency bands were typical N1 and P2 topographies occurred. In particular, this latency band was defined as the time interval between the first occurrence of the N1 topography and the last manifestation of the P2 topography in the segmentation of the microstate analysis.

In order to evaluate any statistical differences in microstates duration and maximal GFP between groups, conditions, and runs, we labelled each time point of the single subject's ERPs according to the map which spatially best correlated with the template maps occurring in the time window of interest. It is important to mention again, that for both the clustering- and the refitting-processes, we adopted the temporal criterion of at least ten time points; thus, a microstate had to persist for at least 20 ms. The statistical evaluation between the two groups was performed by using ANOVAs and based on the following independent variables: duration and maximal GFP of the microstates of interest.

4.1.8 Results

4.1.8.1 Behavioural data

A general linear mixed model analysis for accuracy data and a mixed-effect regression model analysis for RT data were conducted by using *run* (run 1, 2, 3, 4), *condition* (ENV, FS), and *group* (musician, non-musician) as fixed factors, and *participants* as well as *items* as random factors. All post-hoc t-tests were corrected for multiple comparisons by using the Bonferroni procedure.

4.1.8.1.1 Accuracy

The generalized linear mixed model analysis for accuracy data revealed a significant effect of *condition* ($z=8.604$; $p < .001$), which became manifest in terms of a higher accuracy for the identification of the FS stimuli, irrespective of group. Furthermore, we found a significant effect of *run* ($z=4.802$; $p < .001$), which was associated with an increase in accuracy for both groups over the four runs, irrespective of condition. Furthermore, we revealed a significant *condition by run* interaction effect ($z=2.078$; $p =.038$. The *condition by run* interaction was manifested by a stronger increase in accuracy over the four runs in the FS condition compared to the ENV condition. Most interestingly, we found support for our working hypothesis in that the generalized linear mixed model analysis yielded a *condition by group by run* interaction ($z=3.531$; $p < .001$). In order to disentangle this three-way interaction, we performed two separate generalized linear mixed model analyses, one for each condition.

These further statistical analyses yielded a significant *group by run* interaction, but only in the FS condition (FS, $z=4.099$; $p < .001$, ENV, $z=1.320$; $p = .187$). Post-hoc t-tests (two-tailed) calculated for each run across the two groups in the FS condition, clearly showed that the behavioural performance only differed between the two groups during the second run; the musicians exhibited enhanced levels of accuracy as compared to the non-musicians (run 1, $T_{(48)} = -.277$; $p = .783$; run 2, $T_{(48)} = -.2.670$; $p = .010$; run 3, $T_{(48)} = 1.533$; $p = .127$; run 4, $T_{(48)} = 1.284$; $p = .205$).

In summary, the evaluation of accuracy revealed that 1) both groups showed a better behavioural performance during the FS condition; nevertheless, 2) the musicians' performance improved quicker that the non-musicians' performance during the FS condition. This means that the musicians' accuracy increased from the first to the second run faster than the controls' accuracy did.

4.1.8.1.2 Reaction time

The mixed-effect model analysis for reaction time (RT) yielded significant effects for *condition* ($F_{(1,\ 5702)}$ = 296.51; p < .001) and *run* ($F_{(3,\ 5603)}$ = 307.50; p < .001). Furthermore, the same statistical approach revealed a significant *condition by group* effect ($F_{(1,\ 5702)}$ = 18.289; p < .001), as well as a *run by group* interaction ($F_{(3,\ 5603)}$ = 28.928; p <.001). As visible in Figure 1, both groups showed a perceptual learning effect during the course of the experiment that was characterized by a shorter reaction time. In line with the behavioural data resulting from the evaluation of accuracy, all participants experienced more difficulties when performing the ENV condition, which was associated with a longer reaction time.

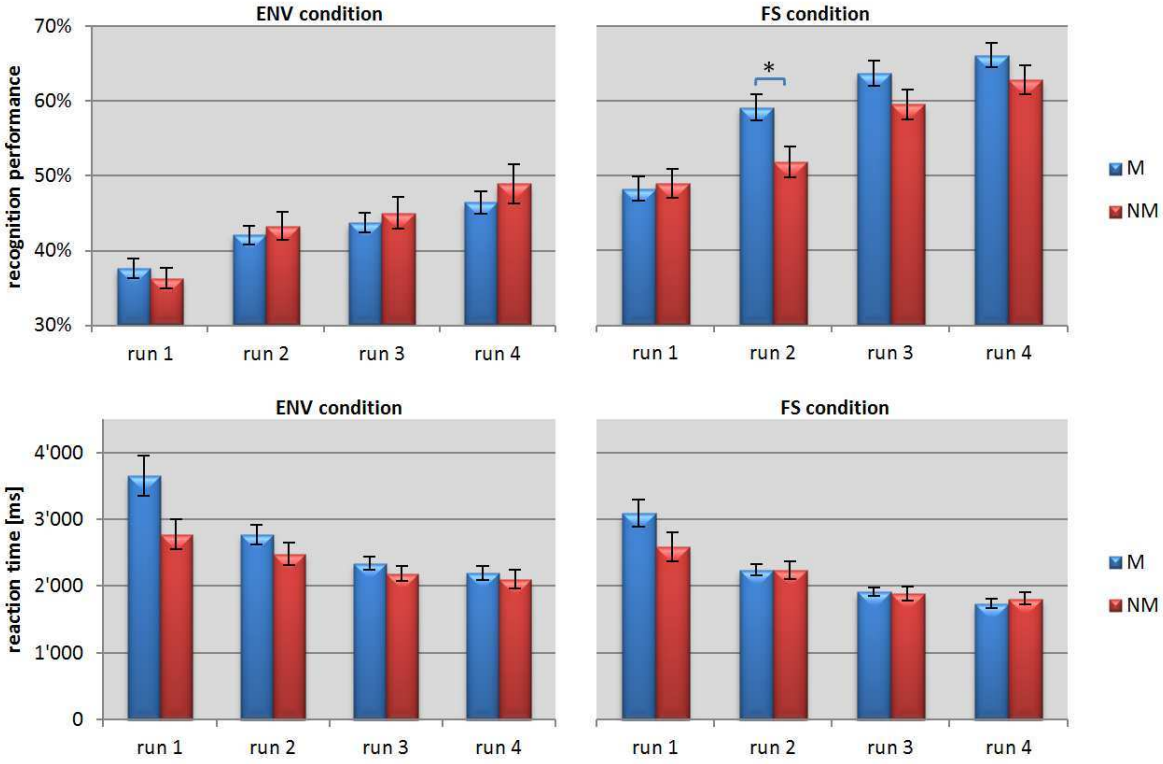

Figure 1: Whereas the top left panel shows recognition performance in the ENV condition over the four runs for both groups, the top right panel depicts recognition performance in the FS condition. The bottom left panel displays the reaction time over all four runs and both groups during the ENV condition. The bottom right panel indicates the reaction time during the FS condition. Musicians are depicted in blue; non-musicians in red.

4.1.8.2 EEG data

4.1.8.2.1 Global field power (GFP)

In order to determine the time windows of interest for the microstate analyses, we computed GFP values for each participant by normalizing data level to the mean GFP for each subject. Group comparisons, which were evaluated by calculating t-tests across the two groups, revealed a significant increase in GFP amplitude in the musician group for the latency band that ranged from 146 to 168 ms after stimulus onset (Figure 2). This latency band overlaps with the temporal dynamic of the auditory evoked N1-P2 complex. Finally, we performed the successive analysis (microstates) by focussing on the N1- and P2-like microstates.

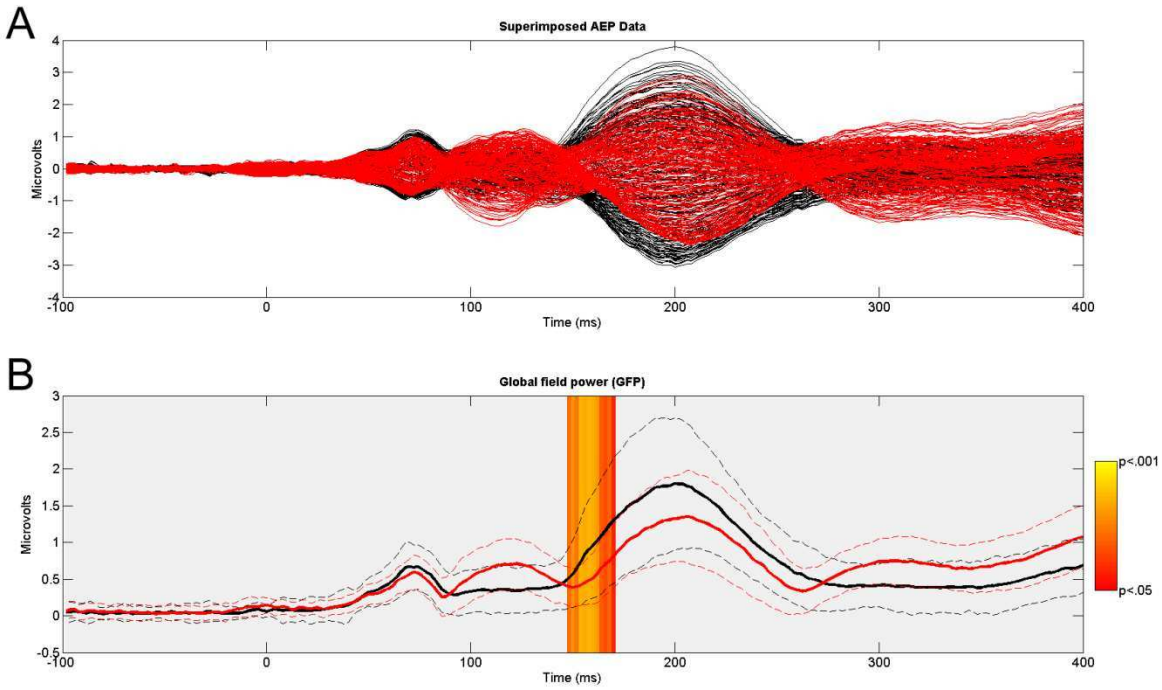

Figure 2: A.) This figure shows separate superimposed AEPs for each of the 204 channels for musicians (black) and non-musicians (red). All conditions and runs were collapsed together. B.) Here GFP related to the musicians (black) and non-musicians (red) are displayed. The dashed lines indicate standard deviation. Significant GFP differences between the two groups are depicted in hot colours. The bar on the right side of image B depicts the significance threshold ranging from .05 to .001.

4.1.8.3 Topographic pattern analyses: microstates

The microstate analyses that we performed on the averaged AEPs of the two conditions (FS, ENV), the four runs (run 1, 2, 3, 4), and the two groups yielded seven representative topographic scalp voltage maps ("template maps"). Figure 3 displays these seven template maps with the associated GFP time courses. Due to the fact that we were specifically interested in the evaluation of the microstates overlapping with the time course of the N1-P2 complex, we restricted our analysis to N1- and P2-like topographic maps at around 100 and 200 ms. We then identified "template map 4" as reflecting an N1-like microstate and "template map 5" as reflecting a P2-like microstate. When considering both the temporal latency of the two maps, as well as the topographic characteristics, the reasoning behind the former step of analysis becomes logical.

In the next processing step, we manually defined the exact time window for the microstate analyses based on the time period between the onset of the N1-like microstate and the offset of the P2-like microstate; namely, from 84 to 380 ms. Next, we refitted all template maps appearing within this time period back to the single subjects` data; that is, template maps 4, 5, 6, and 7. This specific procedure was performed, in order to obtain the duration and maximal GFP value of the topographic "template maps" associated with the N1 and P2 components for each subject, condition, and run. Statistical analyses were performed by computing four separate 4 x 2 x 2 ANOVAs for the following variables: duration, maximal GFP, N1-like microstates, and P2-like microstates. The factors *run* (run1, 2, 3, 4) and *condition* (ENV, FS) served as the within-subject factors, the factor *group* (musician, non-musician) functioned as the between-subject factor.

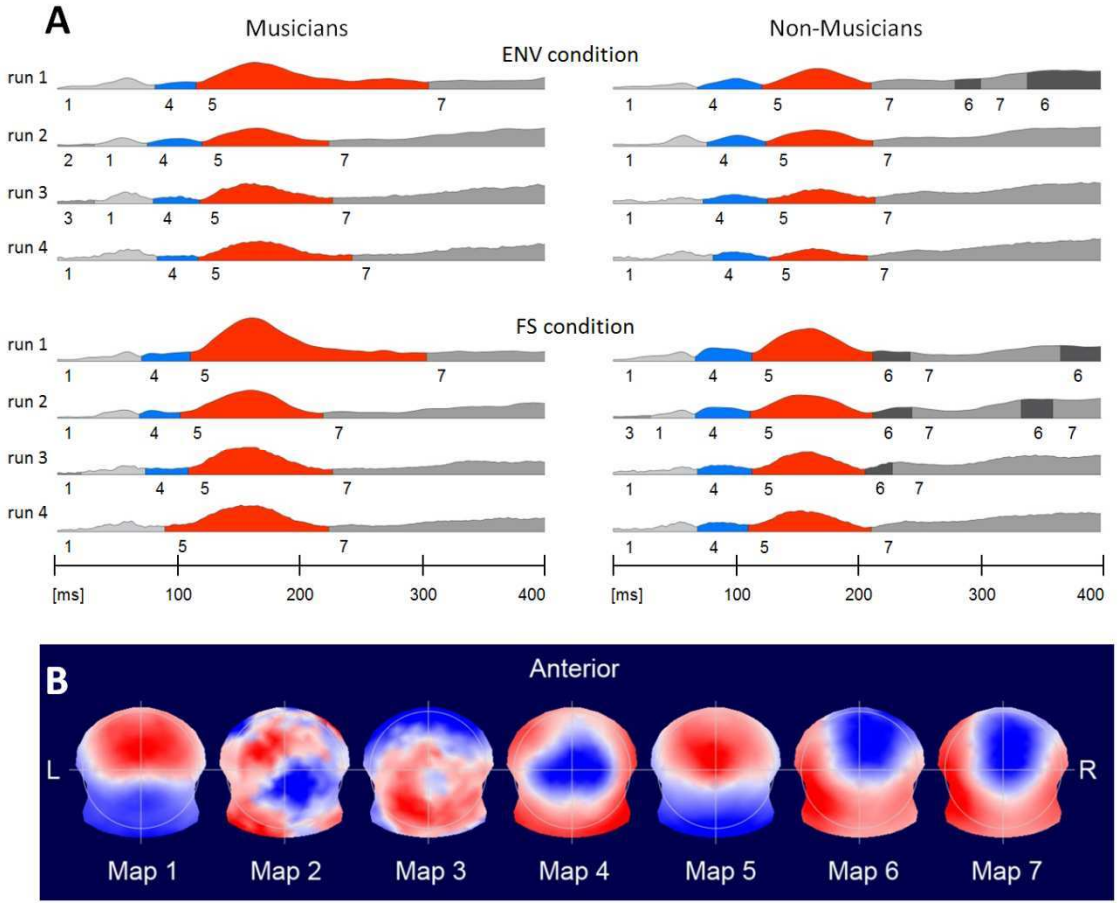

Figure 3: This figure depicts the results of the microstate segmentation procedure. Seven different microstates were found to be most suitable for explaining the data. These seven topographic maps resulted from the microstate analyses that were calculated for the averaged AEPs in the two conditions (FS, ENV), four runs (run 1, 2, 3, 4) and two groups. Template Maps 4 and 5 represent the N1-like- and P2-like microstates, respectively.

Duration of the N1-like microstate

Concerning the duration of the N1-like microstate, the 4 x 2 x 2 ANOVA (repeated measures) revealed a significant main effect for *group* ($F_{(1, 48)} = 121.276$; $p < .001$; $\eta^2 = .158$), which was associated with a significant shorter duration of the N1-like microstate in the musician group. No further significant effects were found in this statistical analysis.

Duration of the P2-like microstate

For the duration of the P2-like microstate, the 4 x 2 x 2 ANOVA revealed a significant main effect for *group* ($F_{(1, 48)} = 10.919$; $p = .002$; $\eta^2 = .185$) that was characterized by a significant longer duration of the P2-like microstate in the musicians, in comparison to the non-musicians. Furthermore, the same analysis revealed a main effect for *condition* ($F_{(1,48)} = 13.889$; $p < .001$; $\eta^2 = .224$). This main effect was associated with a longer duration of the P2-like microstate during the FS condition (Fig. 3). A slight trend was also observed for the factor *run* ($F_{(3,144)} = 3.088$; $p = .060$; $\eta^2 = .060$). Figure 4 depicts this trend as a decrease in the P2-like microstate duration during the time course of the experiment.

Upon examination of Figure 4 it becomes apparent that the P2-like microstate duration fundamentally decreases from run 1 to run 2, at least in the musician group. During the next processing step, we compared the duration of the P2-like microstates across the first two runs and the two groups. In particular, we performed a 2 x 2 x 2 ANOVA analysis with *run* (run 1, 2) and *condition* (ENV, FS) as the within-subject factors and *group* (musicians, non-musicians) as the between-subject factor. This analysis yielded a significant main effect for *run* ($F_{(1, 48)} = 11.542$; $p < .001$; $\eta^2 = .194$), which was characterised by a longer P2-like microstate duration in the first run as compared to the second run, irrespective of group. The same statistical analysis also revealed a significant main effect for *group* ($F_{(1, 48)} = 6.494$; $p = .014$; $\eta^2 = .119$) that was associated with a longer P2-like microstate duration in the musician group in comparison to the non-musicians. We also found a significant *run by group* interaction ($F_{(1, 48)} = 4.815$; $p = .033$; $\eta^2 = .091$). A post-hoc t-test revealed a significant longer duration of the P2-like microstates in run 1 than in run 2, but only within the musician group (musicians, $T_{(24)} = 3.710$; $p < .001$; non-musicians, $T_{(24)} = 0.915$; $p = .369$) .In addition, musicians showed a significant longer duration of the P2-like microstates in run 1 in comparison to the non-musicians ($T_{(24)} = 3.292$; $p = .003$). By contrast, we did not reveal group differences in run 2 ($T_{(24)} = 1.370$; $p = .183$).

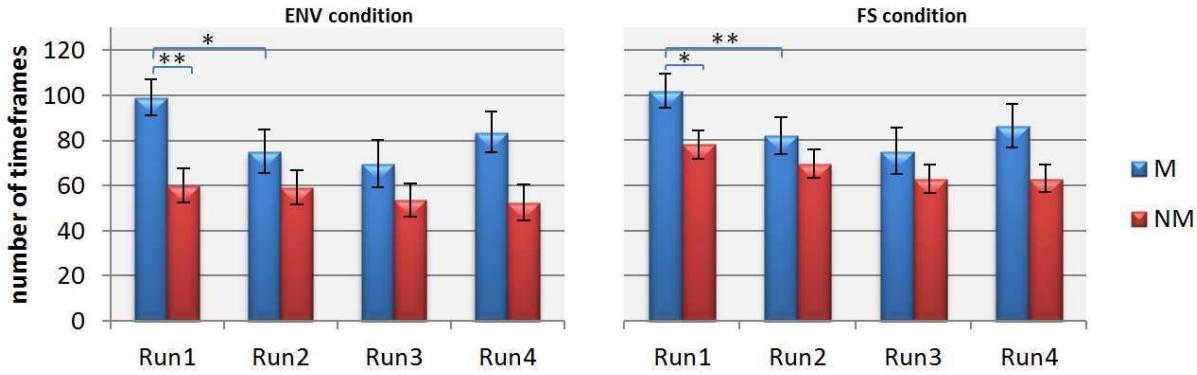

Figure 4: This figure shows the duration of the P2-like microstate for each of the four runs and for each of the two groups. The left panel indicates the duration of the P2-like microstate during the ENV condition. The right panel illustrates the length of the P2-like microstate during the FS condition. Musicians are denoted by blue: non-musicians by red.

Maximal GFP amplitude related to the N1-like microstate

We computed a 4 x 2 x 2 ANOVA (4 runs, 2 conditions, and 2 groups; repeated measures), in order to evaluate the maximal GFP amplitude related to the N1-like microstate. This statistical approach yielded a significant main effect for *group* ($F_{(1, 48)}$ = 5.984; p = .018; η^2 = .111) that was characterized by increased maximal GFP values in the non-musician group. All other main effects or interaction effects did not reach significance.

Maximal GFP amplitude related to the P2-like microstate

Upon evaluation of the maximal GFP amplitude related to the P2-like microstate, (4 x 2 x 2 ANOVA) our analysis revealed a significant main effect for *condition* ($F_{(1, 48)}$= 125.933; p < .001; η^2 = .724). The FS condition elicited increased maximal GFP values relative to the ENV condition (See Figure 5). The same analysis resulted in a main effect for *run* ($F_{(3, 144)}$= 6.948; p = .002; η^2 = .126), which was characterized by a reduction in the maximal GFP during the time course of the experiment in both groups.

Additionally, we performed a less restricted statistical analysis, in order to evaluate whether perceptual learning effects affected the maximal GFP in the two groups

during run 1 and run 2 differently. With this purpose in mind, we evaluated the maximal GFP values related to the P2-like microstate by performing a 2 x 2 x 2 ANOVA with the factors *run* (run 1, 2) and *condition* (ENV, FS) as the within-subject factors and *group* (musicians, non-musicians) as the between-subject factor. We found a significant main effect for *run* ($F_{(1, 48)} = 20.646$; $p < .001$; $\eta^2 = .301$), as well as a statistical trend for the *run by group* interaction ($F_{(1, 48)} = 2.981$; $p = .091$; $\eta^2 = .058$). Post-hoc t-tests revealed a significant decrease in the maximal GFP from run 1 to run 2 only in the musician group (musicians, $T_{(24)} = -4.223$; $p < .001$; non-musicians, $T_{(24)} = -2.013$; $p = .055$).

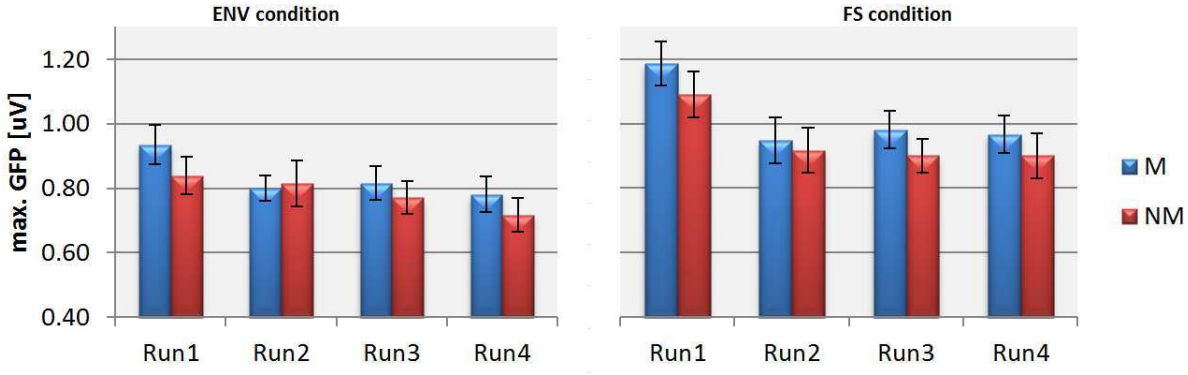

Figure 5: This figure reveals the maximal Global Field Power (GFP) associated with the P2-like microstate for each of the four runs and the two groups. The left panel depicts the maximal GFP of the ENV condition. The right panel illustrates the maximal GFP of the FS condition. Musicians are depicted in blue; non-musicians in red.

4.1.9 Discussion

4.1.9.1 General discussion

Due to previous work showing that professional music training favours segmental (Besson, et al., 2011; Elmer, et al., 2011; Marie, et al., 2011), as well as supra-segmental aspects of speech processing (Moreno, et al., 2009; Oechslin, et al., 2010), we designed an EEG study to specifically investigate the influence of musical expertise on perceptual learning mechanisms. With this purpose in mind, we presented phonotactically legal pseudoword-chimeras to musically trained and untrained individuals. We created these stimuli by independently manipulating the

fine structure and the envelope of the verbal stimuli. Based on converging results that professional music training improves the perception of spectral information in speech (Kraus, et al., 2008; Kuriki, Kanda, & Hirata, 2006; Meyer, et al., 2006; Pantev, et al., 2001b) and music sounds (Magne, Schon, & Besson, 2006), we hypothesised that musicians will recognize an increased number of fine structure manipulated stimuli in the FS condition, compared to non-musicians. We also expected to find that long-term musical training facilitates the recognition and categorization of speech stimuli that are characterized by a reduced envelope structure. This can be expected even though there is only meagre evidence pointing in this direction. In terms of electrophysiological measurements, we expected to find evidence for microstate-differences. Specifically, we hypothesized that there would be differences between the two groups in the fit between latency bands and the auditory evoked N1-P2 complex. (I am not sure if I have understood this sentence correctly) We expected to observe these group-differences during the first half of the experimental session at least. In fact, it is logical to reason that musical experts, who are specifically trained to extract spectral- and envelope-related acoustic information from an auditory stream (for an overview consider Kraus et al. (2009)), will show faster perceptual learning effects than individuals who experience more difficulties when extracting similar information (non-musicians).

4.1.9.2 Behavioural data

During the EEG experiment participants were instructed to assign auditory stimuli, which were characterized by different acoustic manipulations, to one of four pseudowords that were visually presented on a monitor. This sort of task implies that the participants perceive the auditory stimuli and categorize them. The evaluation of the behavioural data showed that all participants increased their recognition performance and reaction time over the four runs of the experiment: this result clearly shows perceptual learning effects in both groups (Figure 1). Our results revealed a main effect for condition (both accuracy and reaction time), which was both irrespective of group and was manifested by an increased accuracy and shorter reaction time in the FS condition as compared to the ENV condition. This evidences

demonstrates that both groups experienced more difficulties performing the ENV condition.

Furthermore, the evaluation of both reaction time and accuracy revealed significant interaction effects. Whereas the evaluation of RT yielded a significant condition by group interaction effect, the analysis of the accuracy data revealed a condition by run by group interaction. A further statistical evaluation of the three-way interaction effect (i.e., condition by run by group interaction) showed a significant run by group interaction effect, but only during the FS condition. Taken together, these results indicate that both groups generally showed a faster reaction time in response to the FS stimuli, and that the musicians' accuracy scores were significantly higher than those of the non-musicians during the FS condition. These results are in line with previous work showing the profound influence of musical training on the processing and detection of spectral modulations in speech (Shannon, et al., 1995; Smith, et al., 2002)) and non-speech signals (Baumann, et al., 2008; Kuriki, et al., 2006; Meyer, et al., 2011a; Pantev, et al., 2001a; Shahin, Roberts, Pantev, Trainor, & Ross, 2005). Our results are novel because we provide evidence for the notion that musical training facilitates the perception and therefore the recognition of pseudowords, which were manipulated in terms of spectral information and do not include envelope-related information. This result may support the view that musicians rely on spectral information more strongly than non-musicians for decoding speech signals. This may be associated with the fact that it is principally the spectral-, and not envelope-related information, which forms the basic foundation of music training.

This study has produced evidence of the existence of perceptual learning effects in the musician group during the FS condition. This effect was manifested by an increase in accuracy from the first to the second run of the experiment in only the musician group. This suggests that musical training favours the perceptual learning of pseudoword-chimeras characterized by spectral information. This specific result supplements the previously described advantage that musicians have in processing spectral speech information. In fact, we not only provide evidence for the fact that musicians are more accurate in recognizing spectral information in speech signals, but also that they learn faster than non-musicians. This result is highly relevant as it

provides evidence for transfer effects from musical training to the perception of spectral information in speech signals (Schneider et al., 2005b).

4.1.9.3 Electrophysiological data

4.1.9.3.1 The effect of experimental conditions

The electrophysiological data are in line with our aforementioned behavioural results. We revealed a main effect of condition, but only in latency bands overlapping with the P2-related component. In both groups, the FS condition elicited stronger maximal GFP amplitudes and was associated with a longer duration of the P2-like microstate, in comparison to the stimuli pertaining to the ENV condition. From a psychoacoustic perspective it is obvious that the fundamental difference between the FS and ENV conditions is that the ENV condition does not contain any spectral information. In this context, previous electrophysiological work has shown that spectrally more complex acoustic sounds are generally associated with an increase in the N1-P2-related amplitudes (Meyer, et al., 2006). An earlier EEG study by Shahin and colleagues (Shahin, et al., 2005) evidenced that it was not the N1 amplitude, which reflected the spectral complexity when the temporal envelope was held constant, but rather the P2 amplitude. The Shahin et al., work clearly points to different spectral encoding mechanisms underlying N1 and P2 responses.

Even though our results are in line with previous work showing a particular sensitivity of P2-related latency bands to the processing of spectral information (Baumann, et al., 2008; Kraus, et al., 2009; Magne, et al., 2006; Meyer, et al., 2006; Shahin, Bosnyak, Trainor, & Roberts, 2003; Trainor, Shahin, & Roberts, 2003), we provide additional information regarding the topographical patterns associated with spectral processing. In fact, by applying the topographic pattern analysis proposed by Murray and co-workers (2008), we provide evidence that spectral complexity profoundly influences the maximal GFP amplitude, especially as regards the duration of the P2-like microstate. This result suggests that the duration of the P2-like microstate can be used as an alternative electrophysiological marker for spectral complexity. Thus, a promising approach in the future will be to reassess whether this

specific electrophysiological marker has the potential to be used for differentiating between people who can hear normally and individuals with specific hearing impairments. In fact, it has been proposed that microstates, which vary in topographical, putative ‘‘atoms of thought’’, may represent different types of information processing (Schlegel, Lehmann, Faber, Milz, & Gianotti, 2011).

4.1.9.3.2 The general effect of musical expertise

Our evaluation of the N1-like microstate yielded a significant shorter duration, as well as a decreased maximal GFP, in the musician group as compared to the non-musicians. In addition, the evaluation of the P2-like microstate revealed a longer P2-like microstate duration in only the musician group. All of these electrophysiological results were not influenced by run or condition.

These results are consistent with previous studies showing differential electrical brain responses in latency bands corresponding to the auditory N1 and P2 components between musically skilled and unskilled individuals during speech- (Marie, et al., 2011; Ott, et al., 2011) and non-speech (Pantev, et al., 1998) processing. Certainly, our study cannot be directly compared with those mentioned above because instead of presenting musical items or natural speech sounds to the participants, we presented pseudoword-chimeras. A further novel aspect of our study is that we are the first research group to focus on microstate analyses for investigating speech processing and perceptual learning effects in professional musicians and non-musicians (Ott, et al., 2011).

It is established that N1- and P2-like topographic maps, which are typically characterized by polarity reversal around the mastoid regions, reflect electrical activity originating from auditory-related brain regions (Vaughan & Ritter, 1970). Nonetheless, we chose to evaluate microstates instead of AEPs; consequently, the differential microstates we revealed between the two groups in latency bands overlapping with the N1-P2 complex are interpreted as reflecting a differential tuning of auditory-related cortical fields as a function of musical training. A similar tuning of auditory-related brain regions in response to speech stimuli as a function of expertise

was described by using EEG (Marie, et al., 2011; Ott, et al., 2011). Since, these differential brain responses between the two groups were irrespective of condition and run, microstates results are interpreted as reflecting an altered sensitivity of musicians for processing acoustic information in general (Kraus & Chandrasekaran, 2010), and speech sounds in particular (Besson, et al., 2011; Meyer, Elmer, Baumann, & Jancke, 2007; Shahin, 2011).

To date there is only sparse literature describing the influence of perceptual and cognitive processes on the spatiotemporal dynamics of the N1-P2 complex (Besson, et al., 2011). Therefore, we can only speculate about the intrinsic meaning of longer or shorter microstate durations in conjunction with musical practice and expertise. Otherwise, there is some evidence indicating that top-down processes are often associated with the modulation of the auditory-evoked P2 component (Hillyard, 1981; Picton & Hillyard, 1974; Schwent & Hillyard, 1975). Furthermore, it is suggested that the N1 component is more influenced by the physical attributes of the acoustic signals, than by other cognitive parameters. Thus, it is possible that our results reflect a differential engagement of perceptual and cognitive strategies between the two groups as a function of musical expertise. In particular, it follows that the intensively trained auditory system of musicians enables a faster and more efficient perceptual analysis of the verbal stimuli. This is reflected by a shorter N1-related microstate duration and a reduced maximal GFP amplitude. This specific pattern of brain response may be particularly advantageous for a faster allocation of cognitive resources, which are necessary for performing the task. Although further studies are necessary to obtain a better understanding of the relationship between microstate duration, perception, and cognition; we consider the topographical pattern analysis presented by Murray and colleagues (2008) as a powerful and novel approach for investigating the electrophysiological correlates of musicianship.

4.1.9.3.3 Musical expertise bolsters perceptual learning mechanisms

The electrophysiological evidence we found for faster learning effects in the musician group is in line with the aforementioned behavioural data. This perceptual learning

effect was reflected in the FS condition by a shortening of the P2-like microstate duration. This is of particular interest because it reveals that perceptual learning effects are reflected by the duration of the microstate associated with the P2-component. We found that microstate duration can be used as a marker for acoustic expertise, as well as for evaluating perceptual learning effects. We can only speculate whether this superior perceptual performance is accompanied by a more efficient engagement of top-down resources, which are necessary for accomplishing this task. Even though we are fully aware that further research is necessary for better understanding which perceptual and cognitive processes are exactly reflected by the microstates duration, our electrophysiological data are in line with previous work indicating that perceptual processing is principally reflected by a modulation of the electrophysiological responses overlapping with the event-related P2 processing stage (Alain, Campeanu, & Tremblay, 2010; Ben-David, Campeanu, Tremblay, & Alain, 2010).

Our experiment revealed a differential modulation of microstates between the two groups in latency bands overlapping with the time course of the P2 component and is assumed to be associated with increased perceptual skills as a function of musicianship. In fact, the musician's increased behavioural performance during the first two runs was accompanied by a reduction of the P2-like microstate duration, as well as a reduction of its maximal GFP amplitude. We observed a decrease of the P2-like microstate duration during the entire experiment in both groups, which was associated with an increased behavioural performance. This finding supports the hypothesis that the duration of the P2-like microstate can be used as a marker for perceptual learning mechanisms.

Finally, it should be mentioned that in the present study we presented disyllabic pseudowords chimeras to a group of musicians and non-musicians in order to investigate the influence of auditory expertise on perceptual learning mechanisms. Basing on previous evidence showing that the increased auditory acuity of professional musicians is primarily reflected by brain responses overlapping with the time course of the N1-P2 complex (Marie, et al., 2011; Ott, et al., 2011), in the present work we only focused on these latency bands. From a pragmatic point of view, this

means that our results are restricted to the processing of the first syllable of the pseudowords. Otherwise, one should consider that the disyllabic words we used were composed of two elements with the same physical modulation, which is spectral or temporal. This implies that the analysis of the first syllable is representative for the processing of perceptual stimulus attributes. Furthermore, in the present work we consciously decided to present phonotactically legal disyllabic pseudowords instead of syllables because the former are indeed more likely similar to real speech. Since the present work primarily focused on perceptual and not cognitive aspects of expertise, we were not interested in evaluating late processing stages. Certainly, further studies are necessary for better elucidate the relationship between microstate duration and different learning mechanisms.

4.1.10 Limitations

The present study was designed to investigate the effect of musical expertise on the perceptual learning of acoustically-reduced pseudoword-chimeras. Even though this study provides novel insights, a few limitations should be noted. Although we provide evidence that microstates can be used as a marker for musical expertise and learning effects, it is nevertheless difficult to explain the perceptual and cognitive processes reflected by this measure. Further studies will be useful to help illuminate the influence of bottom-up and top-down processes on microstates duration, as well as to distinguish between different learning processes. A further limitation of our work is that the sample of musicians we measured was quite heterogeneous. We cannot exclude the possibility that the singers in our sample, who constituted a quarter of the entire group, may have influenced the data in some direction. Further studies are necessary to determine whether the primary instrument played by a musician may have an influence on transfer effects from musical training to speech processing. Finally, even though this study only evaluated perceptual processes restricted to latency bands overlapping with the N1-P2 complex, which were previously shown to constitute an objective marker of auditory processing, further studies are necessary, in order to acquire a better understanding of the microstates associated with subsequent stages of speech processing.

4.1.11 Acknowledgements

We thank Sarah McCourt Meyer for comments on a previous version of the manuscript. This work was supported by Swiss National Foundation (320030-120661 and 4-62341-05).

4.1.12 Conflict of interest

None declared.

4.2 Second Study – Vowel encoding and musicianship

Title of the original study:

The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an EEG study

Jürg Kühnis, Stefan Elmer, Martin Meyer, Lutz Jäncke

Accepted for publication in Neuropsychologia

4.2.1 Abstract

Here, we applied a multi-feature mismatch negativity (MMN) paradigm in order to systematically investigate the neuronal representation of vowels and temporally manipulated CV syllables in a homogeneous sample of string players and non-musicians. Based on previous work indicating an increased sensitivity of the musicians' auditory system, we expected to find that musically trained subjects will elicit increased MMN amplitudes in response to temporal variations in CV syllables, namely voice-onset time (VOT) and duration. In addition, since different vowels are principally distinguished by means of frequency information and musicians are superior in extracting tonal (and thus frequency) information from an acoustic stream, we also expected to provide evidence for an increased auditory representation of vowels in the experts. In line with our hypothesis, we could show that musicians are not only advantaged in the pre-attentive encoding of temporal speech cues, but most notably also of processing vowels. Additional "just noticeable difference" measurements suggested that the musicians' perceptual advantage in encoding speech sounds was more likely driven by the generic constitutional properties of a highly trained auditory system, rather than by its specialization for

speech representations per se. These results shed light on the origin of the often reported advantage of musicians in processing a variety of speech sounds.

4.2.2 Introduction

In the last decades, the investigation of professional musicians has been shown to provide an adequate and natural access for exploring the exceptional faculty of the auditory system to underlie functional (Elmer, et al., 2012; Oechslin, et al., 2012; Ott, et al., 2011) and structural (Hyde, et al., 2009; Jancke, 2009; Moreno, et al., 2009; Munte, et al., 2002) changes. Cortical reorganization processes have previously not only been shown to be restricted to the auditory cortex, but more likely to engage vast parts of cortical (Besson, et al., 2011; Milovanov & Tervaniemi, 2011; Patel, 2011) and subcortical (Kraus & Chandrasekaran, 2010; Kraus, et al., 2008) tissue. Exactly these large-scale neuronal reorganizations, in association with the (partial) overlap between speech and music functions (Patel, 2011), motivated a growing number of scientists to investigate relationships between neuroplastic changes and speech processing in professional musicians (Jäncke, 2012). Until now, some longitudinal studies provide evidence for a causal link between professional musical training and the competence of musicians to actively (Francois, et al., 2012; Fujioka, et al., 2006; Moreno, et al., 2011; Moreno, et al., 2009) or passively (Chobert, et al., 2011) detect small spectral modulations in speech sounds, principally when acoustic cues are difficult to detect (Schön & Francois, 2011). However, a question that has not yet been intensively addressed is whether the cortical reorganizations previously observed in musicians may enable a functional advantage in encoding vowels. In fact, to the best of our knowledge this specific research question has previously only been investigated in children undergoing musical training (Chobert, et al., 2011). The relationship between musicianship and vowel encoding can be expected, than both tones and vowels perception relies on the analysis of spectral cues.

Meanwhile, there is also some evidence from cross-sectional studies showing that professional musicians are advantaged over non-musicians in detecting temporal modulations in CV syllables (or even single words), as provided by previous fMRI

(Elmer, et al., 2012) and EEG (Besson, et al., 2007; Kühnis, et al., 2012; Marie, et al., 2011; Ott, et al., 2011) studies. Interestingly, to date only few EEG studies investigated both spectral and temporal aspects of speech processing within the same sample of musically trained and untrained subjects (Chobert, et al., 2011; Kühnis, et al., 2012). However, these previous studies made use of active rather than passive discrimination paradigms, the former being more profoundly influenced by top-down mechanisms (Brechmann et al., 2007). Hence, the distinctive influence of cognition on auditory functions constitutes a confounding when it comes to compare brain activation patterns between musicians and non-musicians (Baumann, et al., 2008). By contrast, passive auditory paradigms are often considered to constitute a more adequate approach for objectivising bottom-up processes in audition (Naatanen & Alho, 1995).

The faculty to extract and segregate temporal and spectral information from an acoustic stream can be considered a fundamental principle underlying the organization of the auditory system (Bregman, Liao, & Levitan, 1990). This organization principle is not only highly relevant to cope with a complex acoustic environment, but also for facilitating exceptional art forms, like music. Interestingly, this spectro-temporal organization principle is not unique to human beings, than it has previously also been observed in other mammals, like gerbils and rats (Scheich, Brechmann, Brosch, Budinger, & Ohl, 2007). Currently it is still a matter of debate whether the common functioning of the mammalian auditory system may represent a critical evolutionary precursor to lateralized mechanisms for speech processing in humans (Wetzel, Ohl, & Scheich, 2008).

An important step toward the understanding of speech processing was probably the proposition of a division of labour between the auditory cortices of the two hemispheres (Lackner & Teuber, 1973; Poeppel, 2003; Zatorre & Belin, 2001; Zatorre, et al., 1992). According to this framework, it is assumed that the left auditory cortex is more likely involved in processing fast changing acoustic cues, whereas its right counterpart is preferentially driven by slow acoustic modulations. This framework

implies, for example, that the voice-onset time (VOT) (which is a fast changing acoustic cue represented by the time between the release of a stop consonant and the onset of vocal folds vibrations) (Lisker & Abramson, 1964), more strongly relies on left-sided auditory cortex functions (McGettigan & Scott, 2012; Zaehle, et al., 2009). Otherwise, spectral manipulations of the acoustic signal (i.e., long integration windows, fundamental frequency, and formant transitions of vowels) more likely recruit right-sided auditory cortex functions (Abrams, et al., 2008; Johnsrude, et al., 2000; Meyer, 2008). This framework has been supplemented by previous electrophysiological work describing slow and fast spontaneous oscillations originating from auditory-related cortical fields during resting state (Giraud, et al., 2007). Giraud and colleagues provided electrophysiological evidence for the asymmetric sampling in time hypothesis previously described by Poeppel (2003) by showing that right-dominant low-frequency brain oscillations are functionally relevant for integrating acoustic information over long time windows, whereas left-sided high-frequency oscillations are more likely engaged in sampling acoustic signals over short time windows.

Here, we used the EEG technique in order to compare the passive encoding of spectral and temporal speech cues in the auditory system of professional string players and non-musicians while the subjects were passively exposed to spectrally (fundamental frequency and second formant transition of the vowel) and temporally (VOT and duration) manipulated CV syllables. Based on previous EEG studies showing a superior efficiency of the auditory system of musicians in detecting fine-graded spectro-temporal speech (Kühnis, et al., 2012; Nikjeh, et al., 2009; Ott, et al., 2011; Tervaniemi et al., 2009; Zaehle, Jancke, & Meyer, 2007) and non-speech (Koelsch, Schroger, & Tervaniemi, 1999; Tervaniemi, et al., 2005) patterns, we postulated to find increased MMN responses, as well as shorter latencies, in the musicians to all spectral and temporal manipulations. In addition, based on the assumption that the acoustic superiority of musicians is probably driven by the generic constitutional properties of a highly trained auditory system rather than by its specialization for speech representation per se, we expected that musicians will show smaller just

noticeable differences (JND) than non-musicians in response to acoustic modulations in non-speech sounds (pitch, duration, timbre, and gap detection).

4.2.3 Material and methods

4.2.3.1 Participants

Forty-four healthy volunteers (26 female and 18 male) with no past or current neurological, psychiatric, or neuropsychological problems participated in this study. All participants were native Swiss-German or German speakers. The first group consisted of 22 professional string players (13 female and 9 male; primary musical instrument: 9 violinists, 6 violists and 7 cellists; mean age 24.7 years ± 6.8 SD) who commenced their musical training between 2.5 and 8 years (mean age 5.66 years ± 1.44 SD). The musicians we measured practised their musical instrument on average for 2.79 hours per day (min. 0.5 hours, max. 8 hours; ± 1.84 SD). The control group consisted of 22 volunteers without formal musical education (13 female, 9 male; mean age 27.0 years ± 8.2 SD). All participants were paid for participation and gave informed written consent in accordance with the procedures approved by the local ethics committee.

4.2.3.2 Behavioural data

4.2.3.2.1 History of musical training

History of musical training was assessed by using an in-house questionnaire (Elmer, et al., 2012). This questionnaire was adopted to evaluate the age of commencement of musical training, the instruments played, as well as the estimated number of training hours across life span. In particular, the subjects estimated the total number of training hours they performed per day (and per week) in the following periods of life (age): 0–7, 8–10, 11–13, 14–16, etc. Based on these values, we extrapolated the total amount of training hours across life span for each subject.

4.2.3.2.2 Musical aptitudes

The musical aptitudes of the participants were estimated by using the "Advanced Measure of Music Audition" (AMMA) test published by Edwin E. Gordon (1989). This procedure bases on the assumption that to hold music sounds in memory, and to detect melodic and rhythmic variations, constitute a fundamental prerequisite for musical aptitudes. During the AMMA test the volunteers listened to short pairs of piano tone sequences and had to decide whether these sequences were equivalent, rhythmically different, or tonally different. As expected, the musicians performed significantly better than the non-musicians on the AMMA test ($t(42)=20.77$; $p<.001$).

4.2.3.2.3 Cognitive capability

In order to exclude between group differences in cognitive abilities, we applied two short intelligence tests, namely the KAI (Kurztest der aktuellen geistigen Leistungsfähigkeit, (Lehrl, et al., 1991) and the MWT (Mehrfachwahl-Wortschatz-Intelligenz, Lehrl 1977). We did not reveal significant differences between the two groups, this result indicating similar cognitive abilities in musicians and non-musicians. Table 1 provides an overview of the biographical and behavioural data of the participants.

Table 1: Biographical and behavioural data of the two groups (M = musician, NM = non-musicians, AMMA = Advanced Measures of Music Aptitude, KAI = Kurztest der aktuellen geistigen Leistungsfähigkeit, MWT = Mehrfachwahl-Wortschatz-Intelligenz). ** Significant difference between the two groups, ($F_{(1, 42)} = 21.48$, $p < .001$).

	Age [a]		**Gender**		**KAI [IQ]**		**MWT-B [IQ]**		**AMMA** [PR]**	
	M	SD	m	f	M	SD	M	SD	M	SD
M	24.7	6.78	13	9	129.3	8.67	115.5	13.18	77.7	9.97
NM	27.0	8.19	13	9	125.9	9.75	110.2	10.86	56.9	18.51

4.2.3.3 Just noticeable difference

In order to quantify the sensitivity of the auditory system of the participants in response to non-speech sounds, additional offline tests were performed by using an adaptive algorithm (Phillips & Hall, 2002). In particular, after performing the EEG session, each participant underwent four different experimental sessions consisting in detecting the following just noticeable differences (JND): 1) gap detection task with noise bursts (centre frequency of 1 kHz, starting from 100 ms gap duration difference); 2) pitch discrimination task with sine wave tones (1 kHz, starting with 100 Hz difference); 3) duration discrimination task with sine wave tones (1 kHz, 400 ms duration, starting with 100 ms difference); and 4) timbre discrimination task with

a fully artificial vowel (/a/, f0 = 140 Hz, F1 = 680 Hz, F2 = 1150 Hz, starting with a F2 difference of 150 Hz).

Each JND condition consisted of detecting a target stimulus by pressing one out of two response buttons. In particular, in the gap condition the participants had to detect the stimulus with a gap (first or second stimulus), in the pitch condition the higher tone, in the duration condition the longer tone, and in the timbre condition the stimulus with a higher F2 formant transition. In each of the four conditions, the dependent variable was the JND between the two signals. The interstimulus interval was of 1 second, and the next trial started automatically after response selection. Within each condition, the stimulus pairs were randomized by the Presentation Software (www.neurobs.com; version 14.5). Each value of the independent variable was tested by using a two-down/one-up method (Levitt, 1971). Briefly, the starting trials were easy to detect for each subject. After every two successive correct responses, the acoustic manipulation was decreased by a factor of 1.2 in each experimental condition. In case of an incorrect response, the size of the acoustic manipulation was increased by the same factor. This two-down/one-up procedure continued until the subjects performed three errors. To estimate perceptual threshold we averaged the values associated with the three reversal points. Importantly, in order to define robust JND values, this procedure was repeated three times and an average value was calculated. A more precise description of the algorithm we used in the present study was extensively described by Phillips and colleagues (Phillips, Taylor, Hall, Carr, & Mossop, 1997).

4.2.3.4 Stimulus material

In the present study, we used a semi-artificial German CV syllable /ta/ which was created by using the PRAAT software (http://www.fon.hum.uva.nl/praat/). The original syllable was spoken by a professional male speaker and recorded at a sampling rate of 44.1 kHz. In a first processing step, we separated the consonant /t/ from the vowel /a/, and replaced the latter by a fully artificial vowel with identical fundamental frequency (f_0 = 122 Hz) and first two formant transitions (F1 = 680 Hz,

F2 = 1150 Hz) as the original one. In a successive step, we superimposed the pitch and amplitude contours of the original vowel over the artificial one. This procedure is particularly fruitful in that it enables to produce a semi-artificial CV syllable with fully controlled physical features, such as duration, pitch, timbre, and formant transitions of the vowel. During EEG, the experimental manipulation consisted of presenting a standard as well as eight different deviant CV syllables which were altered in terms of pitch, duration, second formant transition of the vowel (F2), and voice-onset-time (VOT). Furthermore, in order to increase the discriminatory power between the two groups, for each deviant we created two deviant magnitude levels. Table 2 provides an overview of the physical parameters of the auditory stimuli presented during the EEG experiment.

Table 2: Description of all stimuli; the "standard"-/ta/ (top line) had a fundamental frequency (f0) of 122Hz with formants F1 of 680Hz, and F2 of 1150Hz, the VOT had a duration of 93ms, though the whole stimulus had a duration of 363 ms.

	pitch f0 [Hz]	formant F1 [Hz]	formant F2 [Hz]	duration [ms]	VOT [ms]
standard	122	680	1150	363	93
pitch_large	136	680	1150	363	93
pitch_small	129	680	1150	363	93
vowel_large	122	680	1300	363	93
vowel_small	122	680	1225	363	93
duration_large	122	680	1150	250	93
duration_small	122	680	1150	300	93
VOT_large	122	680	1150	340	70
VOT_small	122	680	1150	350	80

4.2.3.5 Experimental procedure

The participants sat in a comfortable chair at a distance of about 75 cm from a 19-inches monitor. During EEG recording, the participants watched a silent movie while the auditory stimuli were presented at a sound pressure level of about 65 dB (Digital Sound Level Meter 329, Voltcraft) by using Sennheiser in-ear HIFI-headphones (Sennheiser, CX-350). We used the so-called multifeature "optimum-1" paradigm previously proposed by Näätänen (2004). This paradigm implies that standard (the syllable /ta/) and deviants (the manipulated syllable /ta/) are presented alternating, and that each deviant differs from the last presented deviant in one physical dimension only (pitch, F2 of the vowel, duration, or VOT).

The standard stimulus was presented 960 times (occurrence of 50 %), whereas each deviant was presented 120 times (each deviant occurred in 6.25 % of the trials).

Furthermore, in order to avoid expectation, the interstimulus interval was jittered between 600 and 900 ms. The whole experiment lasted about 32 minutes. The presentation of the auditory stimuli was controlled by the Presentation software (www.neurobs.com; version 14.5).

4.2.3.6 EEG acquisition, pre-processing, and data analysis

Continuous EEG (32 electrodes + 2 eye channels, provided by Easy Cap) was recorded with a sampling rate of 1000 Hz and a high pass filter (0.1 Hz) by using an EEG-amplifier (Brainproducts, Munich, Germany). The electrodes (sintered silver/silver-chloride electrodes) were located at frontal, temporal, parietal and occipital scalp sites according to the international 10-20 system (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP9, TP7, CP3, CPz, CP4, TP8, TP10, P7, P3, Pz, P4, P8, O1, Oz, O2). The reference electrode was placed on the tip of the nose, and electrode impedance was reduced to <10 kΩ by using electrogel conductant. For all pre-processing steps, we used the Brain Vision Analyser software (Version 2.01, Brainproducts, Munich, Germany).

Data were band-pass filtered 1-30 Hz, and artefacts (i.e., eye movements and blinks) were eliminated by using an independent component analysis (ICA) (Jung, et al., 2000) in association with a semi-automatic raw data inspection. Data were sectioned into segments of 700 ms (from 200 ms pre-stimulus to 500 ms post-stimulus), and a baseline correction relative to the -200 to 0 ms pre-stimulus time period was applied. Auditory evoked potentials (AEP) were calculated by averaging the single segments, separately for standard and deviant stimuli. Furthermore, in order to verify that the two groups did not elicit different brain responses to the standard, we statistically compared the AEPs between the two groups at frontal (ROI: FC3, FCz, FC4, C3, Cz, C4) and central (ROI: FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4) regions of interest (ROI). This procedure did not yield significant group differences at any time frame.

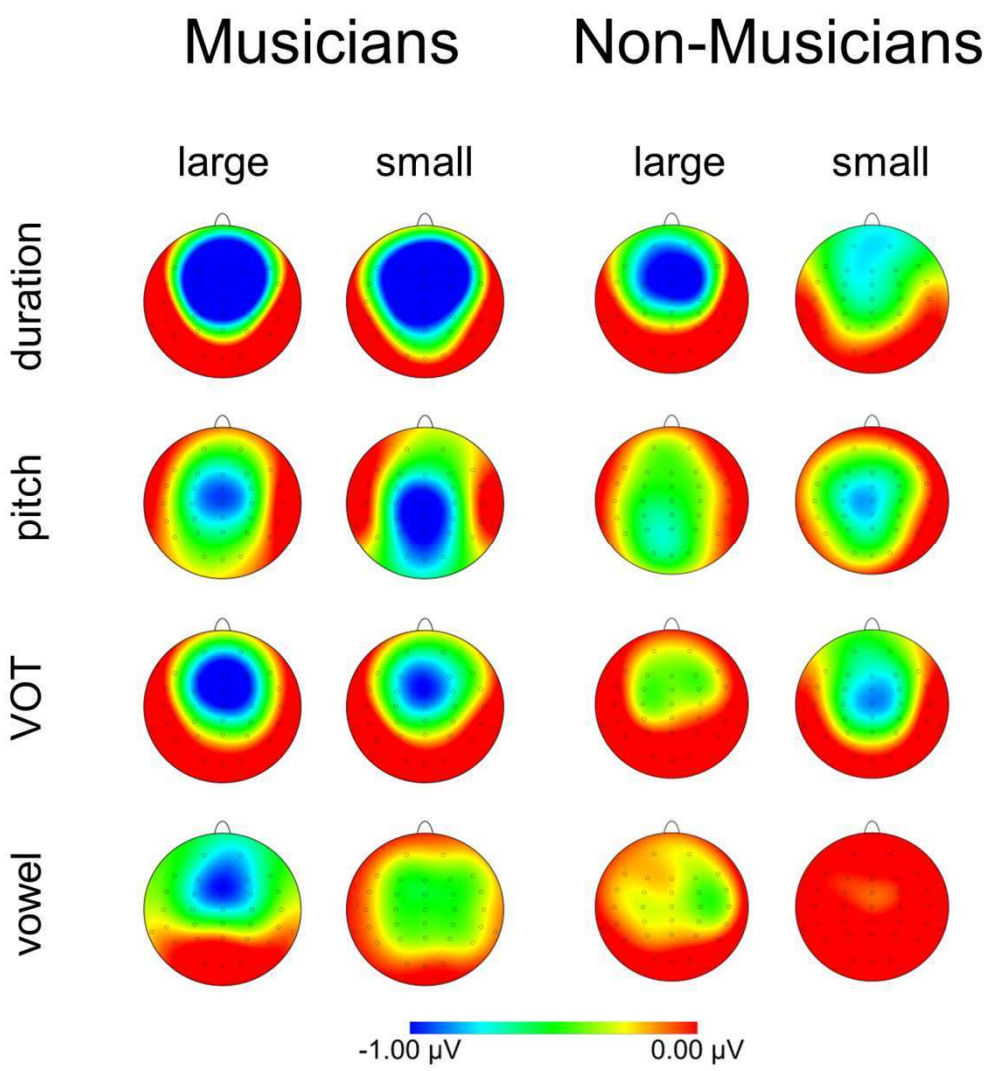

Figure 1: scalp potential distributions of the MMN peaks for both groups (musicians on the left, non-musicians on the right), all four conditions, and the two deviant magnitude levels (large magnitude levels are depicted in the first and third columns, small magnitude levels in the second and fourth columns).

MMN waveforms were calculated by subtracting brain responses to the standards from the deviants. This procedure resulted in eight MMN waveforms. In addition, multi-subject grand averages were computed for each group and condition (see Figure 2). In order to select the electrodes best reflecting the MMN, we computed voltage distribution maps in the time frame showing the maximal MMN amplitude for each group, condition, and magnitude level. Based on these topographic voltage

distribution (see Figure 1), 6 frontal electrodes (F3, Fz, F4, FC3, FCz, FC4) were pooled into one region of interest (ROI). This procedure is particularly useful in that it enables to avoid multiple comparisons between neighbouring electrodes and to increase signal-to-noise ratio (Eichele, Nordby, Rimol, & Hugdahl, 2005; Marie, et al., 2012). In fact, the signal-to-noise ratio increases by averaging the signals of several electrodes. Maximal MMN magnitudes and latencies were labelled semi-automatically at the frontal ROI position, separately for each subject and condition, in a time window of 80 ms around the grand average peak. The labelled peaks were confirmed by visual inspection, and the occurrence of a genuine MMN was tested against zero by using one-sample t-tests. This procedure enables to testify that all deviants elicited a negative deflection reflecting the MMN. In addition, as shown in Figure 2, the presence of a MMN was verified by an inversion of polarity at the mastoid electrodes (TP9 and TP 10). From Figure 2 it appears that during the duration conditions the MMN responses are preceded by a positive deflection. This positivity originates from the temporal asynchrony between the offsets of the standard and deviant stimuli (Jacobsen & Schroger, 2003; Meyer et al., 2011b). The electrophysiological data (MMN amplitudes and latencies) were evaluated by using ANOVAs (repeated measures, 2 groups x 4 deviant conditions x 2 magnitude levels). Figure 2 shows the MMN waveforms (as well as the inversion of polarity at the mastoids) for all conditions, the two groups, and the two magnitude levels.

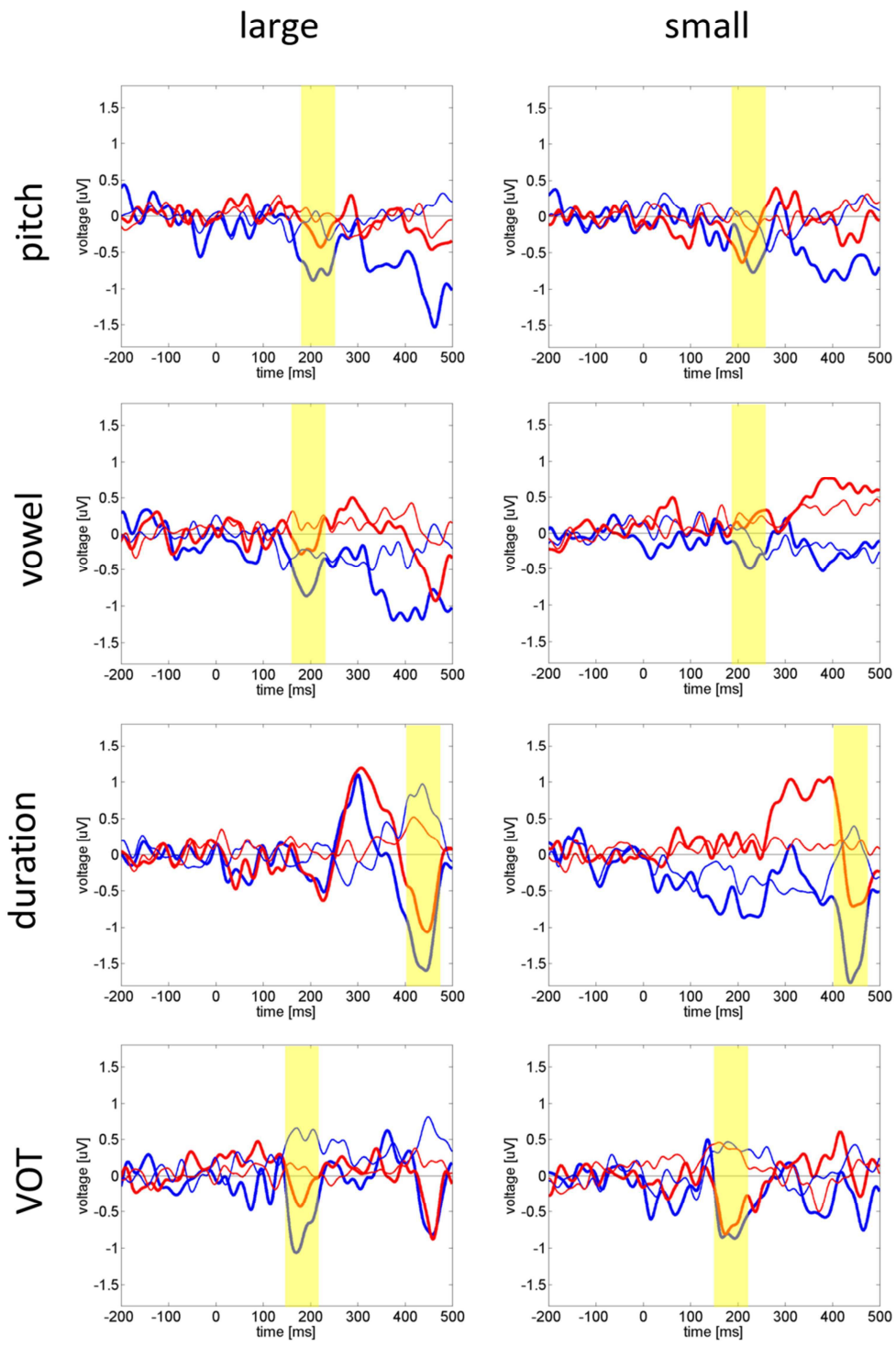

Figure 2: This Figure depicts MMN waveforms and the inversion of polarity at the left mastoid for both groups, all four conditions, and the two magnitude levels. The y-axis depicts microvolt, whereas the x-axis shows time in milliseconds. Musicians = blue, non-musicians = red. Bold lines = MMN; light lines = left mastoid electrode (TP 9).

4.2.3.7 MMN source estimations

In order to get additional information about the putative lateralization of neural sources of the MMNs we used an established source estimation approach, namely LORETA (Pascual-Marqui, et al., 2002; Pascual-Marqui, et al., 1994). It has previously been shown that LORETA permits to approximately estimate the current density maxima of AEP components (Meyer, et al., 2006; Pascual-Marqui, et al., 2002; Pascual-Marqui, et al., 1994), namely the MMN (Gottselig, Brandeis, Hofer-Tinguely, Borbely, & Achermann, 2004; Zaehle, et al., 2009). This approach, unlike conventional dipole fitting, does not require a-priori assumptions about the number and the localization of the dipoles. LORETA calculates the three dimensional distribution of electrically active neuronal generators in the brain as a current density value (A/m^2), and provides a solution for the inverse problem by assuming that the smoothest of all possible activity distributions is the most plausible one for explaining the data. The characteristic feature of this particular inverse solution approach is the low spatial resolution, which conserves the location of maximal activity, but with a certain degree of dispersion (Mulert, et al., 2004). Here we determined the current density distribution for epochs of brain electrical activity on a dense grid of 2394 voxels at 7 mm spatial resolution. The localisation error of LORETA's source identification may vary between 7 (Pascual-Marqui, et al., 1994) and 14 mm (Phillips, et al., 2002). LORETA refers to a three-shell spherical model registered to the Talairach human brain atlas. The source estimations are provided as x, y, z coordinates and are situated relative to the intercommissural line (AC–PC line) in horizontal (x), the anterior/posterior (y), and vertical (z) directions. The solution space is confined to the gray matter portion of the human cortex, which rules out the option that subcortical tissue and white matter contribute to the solution. In the present work, we used LORETA for estimating MMN (peak) sources for all four conditions, separately for the two groups. We used a transformation matrix with high regularization (1e-3*(first eigenvalue), fixed ad-hoc), in order to increase signal to noise ratio.

4.2.4 Results

4.2.4.1 Behavioural results

4.2.4.1.1 Just noticeable differences

Mean threshold values (totally nine measurements) were evaluated by means of a two-way ANOVA (*2 Groups* x *4 Conditions*; repeated measurements). All p-values were adjusted with the Greenhouse-Geisser correction for nonsphericity. This statistical procedure yielded main effects of *Condition* ($F(3,111)=20.14$; $p<.001$) and *Group* ($F(1, 37)=15.15$; $p<.001$), as well as a significant *Group x Condition* ($F(3,111)=3.32$; $p=.026$) interaction. Figure 3 shows the behavioural data of the two groups and the four conditions. Post-hoc t-tests (independent samples, one-tailed, Bonferroni corrected threshold for $p=.05$, $p=.0125$) revealed that the interaction effect originated from a superior performance of the musicians in the gap detection ($t(37)=-4.06$; $p<.001$), pitch discrimination ($t(42)=-4.10$; $p<.001$), and vowel discrimination ($t(40)=-1.80$; $p=.039$) conditions.

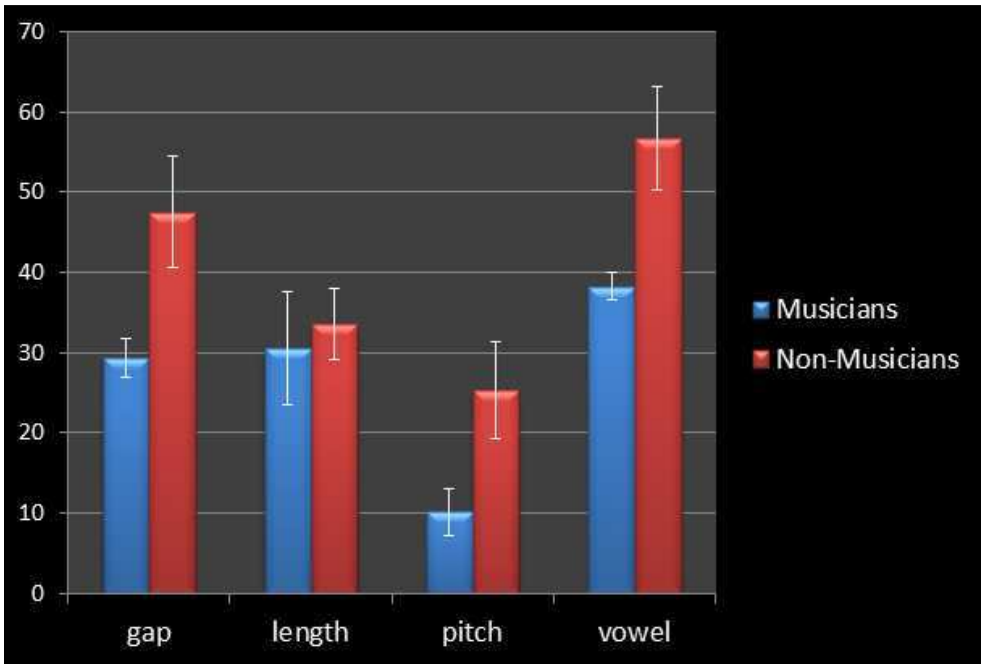

Figure 3: Just noticeable differences (JND) of the musicians (blue) and the non-musicians (red) for all four experimental conditions. For the "gap" and "length" conditions, the y-axis shows the mean JNDs in milliseconds, whereas for the "pitch" and the "vowel" conditions the mean JNDs are reported in Hertz (Hz) on the y-axis.

4.2.4.1.2 Electrophysiological results

As visible from the topographic maps depicted in Figure 1, all experimental conditions elicited genuine MMN responses with a maximal negativity over fronto-central scalp sides, in the time range between 80-180 ms after manipulation of the acoustic signal. Based on these topographic maps, MMN peak amplitudes and latencies were evaluated at the frontal ROI position by means of three-way ANOVAs (repeated measures, 2 *Groups* x 4 *Conditions* x 2 *Magnitude levels*). **The ANOVA** computed on MMN amplitude data revealed a main effect of *Group* ($F(1,42)=13.13$; $p=.001$; $\eta^2=.238$) and *Condition* ($F(3,126)=11.65$; $p=.001$; $\eta^2=.217$. We did not reveal further significant main or interaction effects. As illustrated in Figure 3, the musicians generally elicited significantly increased MMN amplitudes compared to non-musicians irrespective of condition. Since the pitch-small and VOT-small conditions showed relatively small effects, all conditions and magnitude levels were additionally compared between the two groups by performing t-tests for independent samples. This post-hoc analyses yielded significant group differences for all comparisons ($p<.05$), except for the pitch-small ($t(42)=-.224$; $p=.824$), VOT-small ($t(42)=.046$; $p=.964$), and the duration-large ($t(42)=-1.6$; $p=.058$) conditions.

The ANOVA computation executed on latency data (repeated measures, 2 *Groups* x 4 *Conditions* x 2 *Magnitude levels*) yielded significant main effects of *Condition* ($F(3,126)=220.3$; $p<.001$; $\eta^2=.840$) and *Magnitude* ($F(1,42)=35.52$; $p<.001$; $\eta^2=.458$, as well as a *Group x Magnitude* interaction effect ($F(1,42)=4.788$; $p=.034$; $\eta^2=.102$). All data were corrected for stimulus-specific latency shifts. The main effect of condition was driven by prolonged latencies in the *Duration* condition, whereas the main effect of *Magnitude* was associated with longer latencies during the large *Magnitude* level condition ($t(43) = 6.266$; $p< .001$). The significant *Group x Magnitude* interaction effect originated from longer latencies in the musicians in response to the small *Magnitude* levels (mean latency musicians = 151 ms; mean latency non-musicians = 143 ms; $t(42) = 1.834$; $p = .037$). The two groups did not differ in terms of latency in response to the large *Magnitude* levels (mean latency musicians = 191 ms; mean latency non-musicians = 190 ms; $t(42) = 0.299$; $p = .383$).

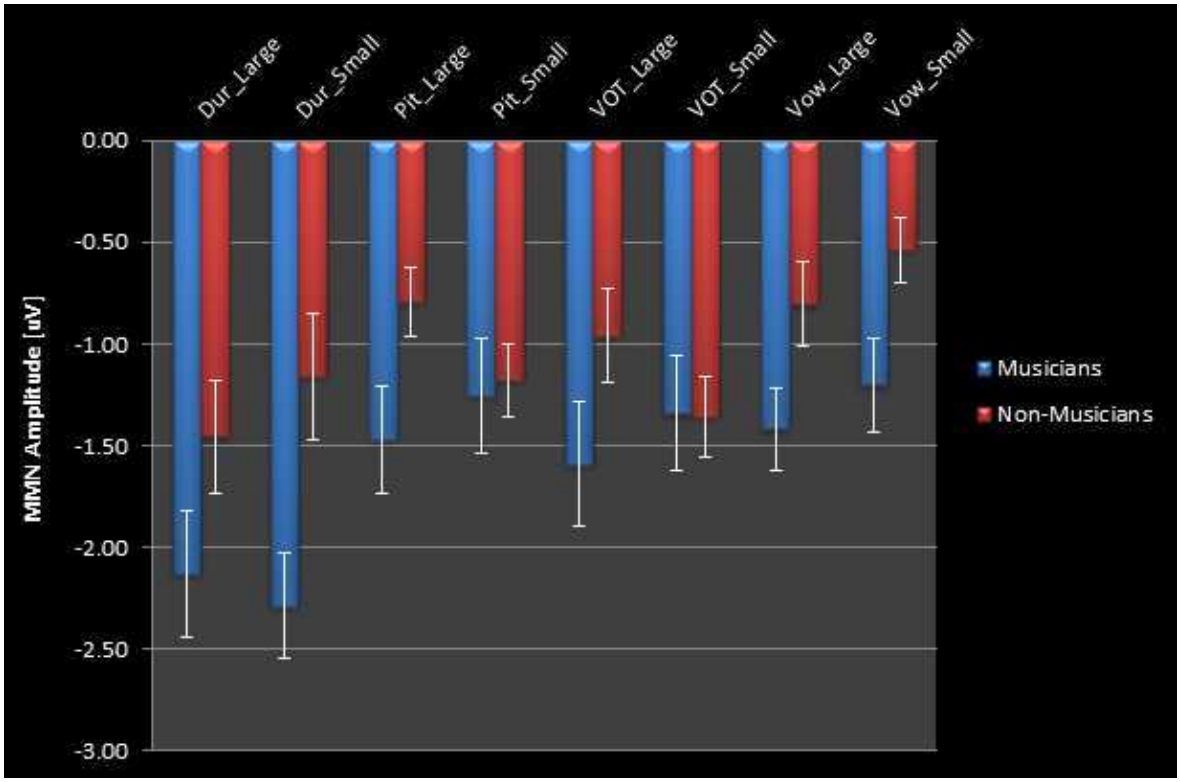

Figure 4: MMN waveform-peaks are depicted separately for the two groups, all four conditions, and both deviant magnitude levels. Musicians are depicted in blue, non-musicians in red. Whiskers illustrate standard errors.

In an additional descriptive analysis we estimated the neuronal sources underlying the MMN responses, separately for the two groups and all experimental conditions. Figure 5 indicates that the temporal manipulations (i.e., VOT and duration) generally more strongly recruited left-sided auditory-related brain regions, whereas the spectral manipulations (pitch and vowel) more likely evoked current density values in the right-sided counterpart. These results are in line with previous work indicating a division of labour between the two auditory-related cortices for the processing of temporal and spectral acoustic features (Poeppel, 2003; Zatorre & Gandour, 2008).

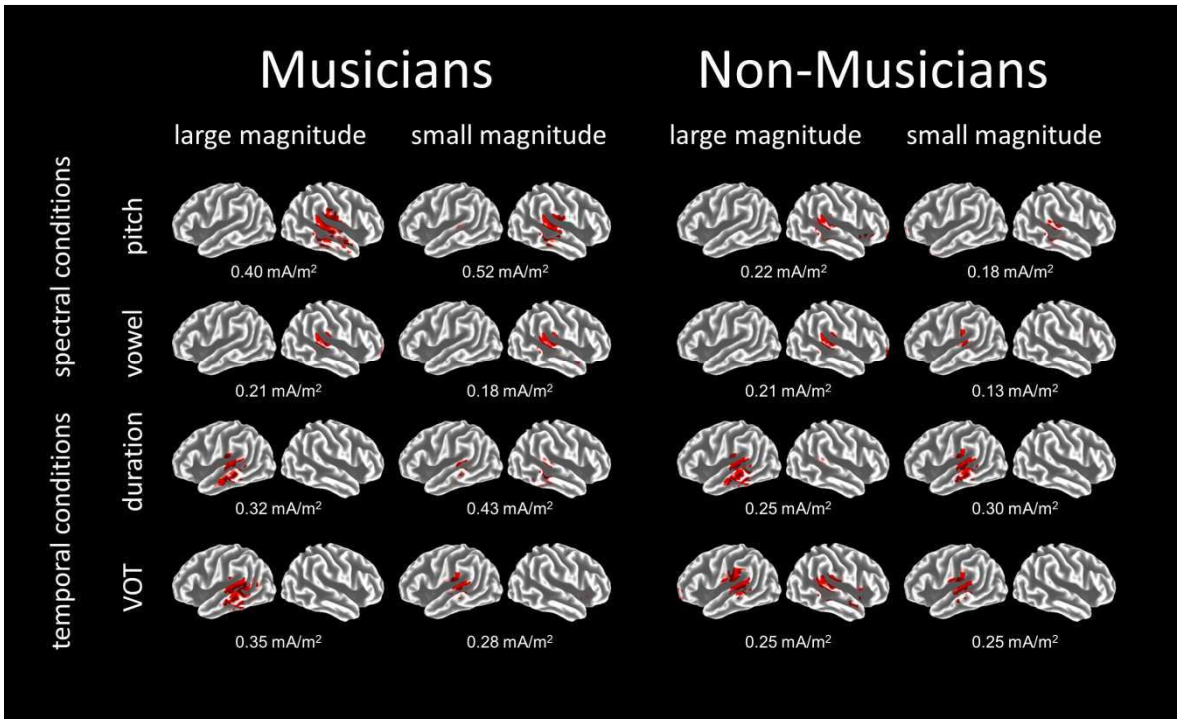

Figure 5: LORETA source estimations of all MMN peaks are depicted separately for the two groups, the four conditions, and the two deviant magnitude levels. In the two spectral conditions (pitch and vowel conditions) a stronger recruitment of the right supratemporal plane can be observed, whereas the temporal conditions (duration and VOT conditions) more strongly recruited left-sided auditory territory.

4.2.5 Discussion

Meanwhile, there is growing evidence pointing to a superiority of professional musicians over non-musicians, in detecting or discriminating spectral (Kühnis, et al., 2012; Nikjeh, et al., 2009) or temporal (Abrams, et al., 2008; Elmer, et al., 2012; Ott, et al., 2011; Zaehle, et al., 2007) variations in speech sounds, or even in both (Tervaniemi, et al., 2009). Although we are fully aware that previous work addressed similar research questions as we did in the present study (Marie, et al., 2012; Ott, et al., 2011; Zaehle, et al., 2009), our experimental design is novel in that it provides important amendments and specifications. First of all, to date only one study investigated both spectral and temporal variations in CV syllables within the same sample of musically trained and untrained subjects (Kühnis, et al., 2012). This is indeed an important approach that permits to hold constant the variance of musical background of the subjects while investigating potential transfer effects from musical

training to the processing of speech sounds. A second important point which distinguishes our work from previous studies is that we used a passive MMN paradigm instead of an active discrimination task. It should be mentioned that even though active listening paradigms enable to conjunctively collect brain- and behavioural responses, they have the shortcoming of top-down driven contaminations. This is problematic, since there is evidence showing that auditory cortical fields and lateralization effects can be strongly modulated by the engagement of task-related cognitive functions (Brechmann, et al., 2007; Jancke et al., 1999; Jancke, Buchanan, Lutz, & Shah, 2001; Jancke, Mirzazade, & Shah, 1999). In order to avoid these experimental confounds and at the same time collect behavioural data that reflect the basic functioning of the auditory system of the subjects, we additionally measured non-verbal JND of spectral and temporal acoustic cues. The reason for presenting non-speech sounds during these offline tests is supported by previous work indicating that the increased cortical representation of speech sounds in musicians is probably more likely related to an overall increased sensitivity of the auditory system for processing subtle acoustic cues rather than to a more robust commitment to speech sounds per se (Kühnis, et al., 2012). A further novelty of our study is that we addressed the question whether musicians may have a perceptual advantage in processing spectrally rich speech sounds, namely vowels. In fact, to the best of our knowledge this has not yet been shown before in adults, but only in children undergoing musical training (Chobert et al. 2011). Next, we shall place the results of the present study in a broader context by separately discussing the neuronal encoding of spectral and temporal CV manipulations.

4.2.5.1 Spectral manipulations of the vowel

The basic idea behind the manipulation of vowels is that if musicians are generally better in perceiving tonal information (Magne, et al., 2006; Schon, Gordon, & Besson, 2005), they should also have a perceptual advantage in encoding vowels, especially in situations where acoustic discrimination is difficult (Elmer, et al., 2012; Kühnis, et al., 2012; Marie, et al., 2011). In fact, different vowels differ in terms of spectral information (i.e., formant transitions) (Liu & Kewley-Port, 2004), and both vowel and

tone perception rely on the analysis of similar (if not the same) spectral cues (Kraus & Chandrasekaran, 2010). If the auditory system of subjects with musical expertise is more efficient in processing spectral information, then musicians and non-musicians should elicit differential MMN responses in those auditory-related cortical fields which are preferentially driven by spectral information. Hence, based on the division of labour between the left and right auditory cortices (McGettigan & Scott, 2012; Poeppel, 2003), we expected to provide evidence for differential MMN responses between the two groups mainly originating from the right-sided supratemporal plane. This reasoning is further supported by some brain imaging and dichotic listening studies which have shown that vowel perception is mediated by the right auditory system in general (Studdert-Kennedy & Shankweiler, 1981), and by the right-sided planum temporale in particular (Jancke, Wustenberg, Scheich, & Heinze, 2002b).

Our findings are fully in line with the above mentioned line of argumentation in that we found significantly increased MMN responses to spectrally manipulated vowels in musical experts, however principally in response to large deviant magnitudes. From an electrophysiological perspective, increased auditory-evoked MMN amplitudes can be explained by the recruitment of additional neuronal ensembles residing in auditory-related brain regions, or even by an enhanced synchronicity of neuronal spikes (Kraus, et al., 2009; Krishnan & Gandour, 2009; Schreiner & Winer, 2007; Tzounopoulos & Kraus, 2009). Unfortunately, to date only a few studies investigated the neuronal encoding of fundamental frequency- and formant manipulations in vowels by measuring musicians (Bidelman & Krishnan, 2010; Parbery-Clark, Tierney, Strait, & Kraus, 2012) or even musically trained children (Chobert, et al., 2011).

Of further interest for our work are recent studies showing that musicians demonstrate much more robust subcortical representations of closely related speech sounds differing in their harmonic spectra only (i.e., the formant trajectories) (Parbery-Clark, et al., 2012; Wong, Skoe, Russo, Dees, & Kraus, 2007). Notably, currently there is no previous evidence showing altered cortical (rather than

subcortical) brain responses to formant manipulations of vowels in adult musicians. Certainly, although our results constitute an important contribution to better understand the neuronal implementation of vowels in musicians, it results difficult to fundamentally explain the origin of this specific effect. In fact, it is conceivable that the increased MMN responses we revealed in the musicians in response to large deviant vowels were driven by bottom-up projection from the brainstem up to auditory cortical fields. Alternatively, it can also not be excluded that the results provided by Parbery-Clark and co-workers (2012) were driven by modulatory corticofugal projections from the auditory cortex to the brainstem. Certainly, the definitive answer to this question can only be provided by further studies dedicated to investigate the dynamic interplay between cortical and subcortical brain regions.

4.2.5.2 Temporal manipulations of CV syllables

A further important question addressed by our work pertains to the relationship between musicianship and the perception of temporal speech cues, namely VOT and duration. The reasoning beyond this relationship implies that the auditory system of musicians is generally more sensitive. This should also have some influences on the perception and processing of temporal cues in speech sounds. Even though this superiority has never been shown before in string players, a handful of work has previously addressed similar research questions by means of active fMRI (Elmer, et al., 2012) and EEG (Kühnis, et al., 2012; Marie, et al., 2011; Ott, et al., 2011) paradigms. Based on previous work showing increased MMN amplitudes in response to temporal speech and non-speech manipulations in children (Chobert, François, Habib, & Besson, 2012; Meyer, et al., 2011b) and adults (Marie, et al., 2012) with musical training, we anticipated to find a similar response pattern in string players. In addition, as previously shown by an fMRI study which adopted similar stimuli as we did in the present work (Elmer, et al., 2012), we expected to find some evidence for a left-lateralized response pattern in response to temporally manipulated speech cues.

Our results are consistent with previous work showing that musicians elicited increased MMN amplitudes in response to changes in duration and rhythmic structure of music sounds (i.e. Russeler, Altenmuller, Nager, Kohlmetz, & Munte, 2001; van Zuijen, Sussman, Winkler, Naatanen, & Tervaniemi, 2005; Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009; Vuust et al., 2005). Our results lead us to conclude that the increased representations of temporal acoustic cues in musicians in response to musical sounds likewise favour the temporal encoding of speech sounds. In this context, it results plausible to assume that musicians benefit from the common processing mechanism underlying both speech and music processing (Kraus & Chandrasekaran, 2010). However, on the other side it remains somewhat difficult to explain how exactly musical training influences the processing of VOT. In fact, whereas duration is an important acoustic cue that enables musicians to follow a particular musical rhythm, transient acoustic signals like the VOT are certainly not likewise specifically trained by string players. In fact, one would rather expect such a superiority in pianists who are regularly passively exposed to explicitly involved in playing rapid, virtuosic, and regular melodies characterized by small temporal gaps between the single notes (Schneider, et al., 2005a). Otherwise, different attack times can be considered as important expressive cues in string players. In fact, even though the reactivity of stringed instrument is much lower than for example that of piano, nevertheless professional string players are highly trained to perceive temporal acoustic cues, which constitute a fundamental requirement for a precise musical timing in an orchestra. Consequently, we may speculate whether the generic (rather than specific) practice of temporal cues (i.e., rhythm, tone duration, etc.) may be sufficient to fundamentally increase the functional capacity of the left auditory-related cortex, a brain region previously shown to support the processing of temporally changing acoustic information (Jamison, Watkins, Bishop, & Matthews, 2006; Meyer et al., 2005; Zaehle, et al., 2007). This reasoning may further imply that all professional musicians can additionally benefit from musical training in that they gain more efficiency in encoding temporal attributes of speech sounds, irrespective of the reactivity of their primary instrument (Elmer, et al., 2012; Kühnis, et al., 2012; Marie, et al., 2011; Ott, et al., 2011). Certainly, further empirical studies investigating the processing of temporal speech sounds in different sample of musicians (pianists,

drummers, string players, singers, etc.) are required for shedding more light on this issue.

Finally, a question that cannot be answer in this study is why musicians showed enlarged MMN magnitudes in response to CV syllables characterized by large VOT manipulations only. Previous studies reported that musicians have a functional and behavioural advantage in processing speech stimuli principally when the differences are difficult to detect (Elmer, et al., 2012; Marie, et al., 2011) Hence, we may speculate whether the homogeneous sample of musicians we measured (namely all string players) may have influenced our results.

It is therefore possible that other kinds of musical practice than string training may have a stronger influence on the functional capacity of the auditory system to detect small temporal variations. In order to give a definitive answer to this question, further studies should measure groups of musicians playing different instruments.

4.2.5.3 The influence of musical expertise on deviant magnitude

In contrast to previous work (Tervaniemi, et al., 2005; Tervaniemi, et al., 2009), we did not reveal a modulation of MMN amplitudes as a function of the magnitude of deviation. Even though we are fully aware that MMNs usually increase with deviation level (Naatanen & Alho, 1995), we think that the absent amplitude modulation as a function of deviant magnitude may result from the physical properties of the stimulus material we used. In fact, by considering the difference values between standard and deviant CV syllables (see Table 2), it becomes obvious that they are rather small, in both magnitude level manipulations. For example, in the pitch condition the deviant CV syllables were manipulated by increasing the fundamental frequency of 5% and 10%, respectively. This implies that the two deviant stimuli of the pitch condition only differed in fundamental frequency by 5%, corresponding approximately to a half-tone. Our line of argumentation is further supported by a previous work of Partanen and colleagues (2011), who presented the participants deviant CV syllables for which f0 were even more strongly manipulated (8% and 15%) than our stimuli.

Nevertheless, Partanen and colleagues also did not provide evidence for differential MMN responses to the deviants varying in magnitude level.

4.2.5.4 Behavioural results

A particular innovative aspect of our work is that we additionally performed auditory threshold measurements (JND) for evaluating the elemental functioning of the auditory system of musicians and non-musicians in response to non-speech sounds. In particular, we tested whether the cortical representation of pitch-, duration-, and timbre-discrimination, as well as of gap detection, may explain the increased MMN responses to speech sounds we observed in musicians. In this context, we provide evidence for an overall and universal superiority of the auditory system of musicians to perceive small acoustic variations. In addition, the significant *Group x Condition* interaction effect revealed that the increased sensitivity of musicians to distinguish non-speech sounds was particularly evident in the pitch (f0) and gap conditions. Hence, our results lead to suggest that the increased MMN responses we observed within the musicians in response to acoustically manipulated CV syllables were principally driven by an increased sensitivity of the auditory system rather than by an optimized speech representation per se. Hence, we would like to suggest that future work dedicated to uncover putative transfer effects in musicians should also include non-verbal threshold measurement. Non-verbal JND measurements are indeed ideal in that they enable a more differentiated reasoning about the primary causes of a more robust encoding of speech sounds in the auditory cortex of musicians.

4.2.5.5 Neural bases of the cortical asymmetry

To date, there is strong evidence that musical training specifically affect functional asymmetry of the auditory system (Oechslin et al. 2010, Ohnishi et al. 2001). For example, Elmer and colleagues (Elmer, et al., 2012) showed that musical training is associated with increased brain responses in the left PT, supporting the idea that this area has been reorganized in musicians which at the end results in different categorization of fast changing acoustic cues. In addition, an EEG study which

investigated musicians and non-musicians (Boh, et al., 2011) revealed a left-hemispheric lateralization of MMN responses while participants passively listened to six-tone patterns. Since the left auditory cortex is especially involved in the processing of rapidly changing acoustic cues (Poeppel 2003; Giraud et al. 2007), the results of Boh and colleagues may indicate that additional networks that help to structure temporal patterns (left hemisphere) might be recruited for processing demanding pitch cues. These findings can be explained in the context of the asymmetric sampling in time framework (Poeppel, 2003) which emphasizes a division of labour between the two hemispheres. Giraud and colleagues (2007) proposed that this division of labour may be driven by different properties of the neuronal cortical ensembles residing in the left and right supratemporal plane. In particular, the authors proposed that high frequency oscillations in the left auditory-related cortex favour the processing of fast changing acoustic patterns, whereas low frequency oscillations originating from right-sided auditory fields are more likely involved in processing slow acoustic modulations. It is reasonable to assume that high-frequency gamma oscillations are functionally more efficient for encoding fast acoustic changes, whereas low frequent fluctuations in the theta frequency-range are more convenient for detecting slow acoustic modulations (Giraud, et al., 2007). This reasoning can be amended by taking into account neuroanatomical models which suggest a different spacing and myelinisation of the columnar organization of the right and left auditory cortices (Rakic, 1995). In this context, one may assume that the increased spacing of cortical columns in the left auditory cortex, in association with a stronger myelinisation, constitutes a highly efficient architecture for detecting and processing fast changing acoustic information. By contrast, the more densely packaged and less myelinated columns residing in the right auditory cortex more likely constitute an optimal anatomical prerequisite for processing slow acoustic modulations. However, to date the exact mechanisms which modulate this specific asymmetry as a function of musical training are still a matter of debate.

Our data are generally in line with the functional asymmetry of the two bilateral auditory cortical fields which are differentially involved in processing slow vs. fast

changing acoustic modulations. In fact, in both groups source estimations generally indicated increased current density clusters in response to temporal cues, namely VOT and duration most likely originating from left-sided perisylvian brain regions. Complementary, the same approach revealed a functional superiority of the right-sided auditory cortical fields for the processing of spectral manipulations (i.e., pitch and vowel). It is important to mention that our results partially diverge from those previously reported by Chobert et al. (2011) who investigated electrophysiological brain responses to CV syllables which were manipulated in the temporal and spectral dimension in a sample of musically trained children. In particular, Chobert and colleagues provided evidence for left-lateralized MMN responses (measured over the scalp without applying source estimation approaches) to duration deviants but not to VOT deviants. By contrast, the source estimation approach we used in the present study generally indicated a left-sided lateralization for the processing of both temporal manipulations (i.e., VOT and duration) in the musicians as well as in the nonmusicians. Even though we are fully aware that the source estimation approach we used in the present work is purely qualitative in nature and does not base on statistical procedures, there are anyhow several possible explanations for the divergence of our results and those of Chobert and co-workers. First of all, it should be considered that Chobert and colleagues measured musically trained children who certainly do not have achieved full brain maturation. This is an important issue, because it has previously been shown that hemispheric lateralization increases during the course of development (Kast, Elmer, Jancke, & Meyer, 2010). Second, the lateralization effect reported by Chobert and colleagues was found for ERPs measured over the scalp. It is actually impossible to infer the underlying neural generators from scalp EEG data due to the “inverse problem”.

4.2.6 Conclusions

By using a multi-feature MMN paradigm we show for the first time that professional string players demonstrate an increased auditory representation of spectrally manipulated vowels as well as temporally manipulated CV syllables. In addition, our results clearly indicate that the increased responsiveness of the musicians’ auditory

cortex to vowels and CV syllables is not at all speech specific. In fact, non-verbal threshold measurements (JNDs) rather support the view that professional musical training generally improves the generic constitutional properties of the auditory system.

4.2.7 Fundings

Swiss National Foundation 320030B_138668, Zürcher Universitätsverein (FAN)

4.2.8 Acknowledgements

We thank Sarah McCourt-Meyer for comments on a previous version of the manuscript. Authors` contributions: J.K. conceived the study, performed the EEG measurements, the statistical analyses, formulated the hypotheses, and drafted this manuscript. S.E was involved in the interpretation of the data, statistical analyses, and in drafting the manuscript. M.M. participated in the study design, study coordination, hypothesis formulation, and contributed to the manuscript. L.J. contributed to the study's hypothesis, design, results, discussion, and was also involved in the preparation of this manuscript. All authors read and approved the final version of the manuscript.

4.2.9 Conflict of interest

None declared.

5 General discussion

In the last section of this thesis, I will discuss the findings of my studies in the context of several open questions.

Both presented studies investigated the influence of musical expertise on speech perception. The first study examined in an active learning task the influence of musical expertise on top-down processes involved in speech perception. This study provided evidence showing that musical expertise strongly influences the perception of speech. Musicians learnt much faster pseudowords which were reduced to their spectral component only. The second study used a passive mismatch negativity paradigm to investigate the influence of musical expertise on bottom-up processes of basal speech perception of CV syllables manipulated in their basic feature, namely in pitch, duration, VOT, and the second formant transition of the vowel. Also in the second study, musicians showed stronger electrophysiological brain responses on small speech aberration.

5.1 Musical expertise and speech perception

As previously shown, musical expertise does not only influence top-down processes of speech perception but also bottom-up processes. A variety of studies have revealed that top-down processes are often associated with intensified P2 component in the AEP (Hillyard, 1981; Picton & Hillyard, 1974). The results of the first study revealed prolonged and increased P2 component while learning pseudowords. Thus, it is possible that these results reflect a differential engagement of perceptual and cognitive strategies between musicians and non-musicians as a function of musical expertise. Otherwise, also bottom-up processes may support these results. The intensively trained auditory system of professional musicians enables a faster and more efficient perceptual analysis of the verbal stimuli. This is reflected by reduced N1 components. This specific pattern of brain response may be particularly advantageous for a faster allocation of cognitive resources, which are necessary for performing the learning task.

In the light of these results, the **first open question**

Do musicians show a perceptual learning advantage in contrast to non-musicians, when they hear words limited in spectral or temporal information only?

must be accepted provisionally. These results are supported by other authors showing enhanced top-down processes for musicians (Besson, et al., 2007; Chobert, et al., 2011; Oechslin, et al., 2012; Tervaniemi, et al., 2009). Nevertheless, in the first study, musicians only performed better in the condition where auditory chimeras contained their verbal information in the fine structure, i.e. the spectral information. This means that musical expertise stronger enhances the analysis in the spectral domain than in the temporal domain; musicians lose their perceptual advantage when missing spectral information. Interestingly, Shahin et al. (2005) demonstrated in an EEG study that the P2 rather than the N1 component encodes spectral complexity when the temporal envelope was held constant. This may support the idea that musical expertise especially advances top-down processes in the spectral domain. This further may explain why musicians outperformed non-musicians in the perceptual learning task of pseudowords reduced to their spectral information.

Otherwise, both studies presented in this thesis provided evidenced for the support of musical expertise on bottom-up processes in speech perception. On the one hand, in the first study, musicians showed reduced N1 components, i.e. shorter microstate and diminished magnitude. Today, it is widely accepted that the N1 component is influenced mainly by the physical attributes of the acoustic signals than by other cognitive parameters, i.e. bottom-up processes (Naatanen & Michie, 1979). On the other hand, the second study showed stronger MMN on different CV manipulation. Both, the N1 as well as the MMN are components stronger characterized by bottom-up processes rather than top-down ones (Naatanen & Alho, 1995; Näätänen, Gaillard, & Mäntysalo, 1978; Naatanen & Michie, 1979). These results suggest that musical expertise supports both, top-down and bottom-up processes of speech perception.

In this context, the **second open question**:

Do musicians have an advantage to better recognize differences in manipulated CV syllables?

must be accepted provisionally. This third question was clearly verified in the light of bottom-up processes characterized by stronger MMN in the second study presented in this thesis. Schneider et al. (2005a) investigated with a combination of structural MRI and MEG a large sample of professional musicians. They showed that those musicians who played an instrument producing short, sharp, or impulsive tones (e.g., drums, guitar, piano, trumpet, or flute) were associated with both larger gray matter volume and enhanced P50m activity in the left auditory cortex, which is sensitive to rapid temporal processing. Otherwise, the musicians who played melodic instruments that produce rather sustained tones with characteristic changes in timbre (e.g., bassoon, saxophone, French horn, violoncello, or organ) exhibited a more dominant right auditory-related cortex, which is known to be sensitive to slower temporal and richer spectral processing. This may support the hypothesis that practicing of a musical instrument specifically fosters neural plasticity in the right and left auditory cortices. The second study showed clearly that musicians elicited stronger preattentive neural responses on either temporally or spectrally manipulated CV syllables than non-musicians. Moreover, this supports the aspect that musical expertise optimizes the neuronal configuration of the auditory cortices what may lead to enhanced basal spectral and temporal analysis of the on-going auditory signal.

5.2 Lateralization

With the AST hypothesis (Poeppel, 2003) in mind, one would expect to find different neural activation pattern while processing temporal and spectral features of speech. Poeppel argue that both auditory cortices are involved in the analysis of acoustic signals. The left auditory cortex processes these signals in short time windows whereas the right auditory cortex processes acoustic signals in longer time frames. In the context of speech perception, this means that the right auditory cortex encodes

slower alteration in speech such as prosody, pitch contour or vowel colours. Otherwise, the left auditory cortex encodes fast variations such as short temporal changes, e.g. VOT.

The second thesis presented here in this thesis, showed on a descriptive level exactly these lateralization effects predicted by Poeppel (2003). Spectrally manipulated CV syllables evoked more intense right sided brain activation. The estimation of the underlying brain generators localized this activation in the right auditory cortex. Otherwise, temporally manipulated CV syllables evoked such brain activation localized stronger in the left auditory cortex. In the light of this observation the **third open question**:

Is it possible in the context of the AST (asymmetric sampling in time hypothesis by Poeppel, 2003) to show any lateralisation effect in processing CV syllables manipulated in temporal and spectral features?

must be accepted provisionally.

It must be remarked at this point, first, the second study did not reveal statistical significant lateralization differences. The results of the invers solution did not survive any statistical analysis, but on a descriptive level a predictable lateralization effect was clearly visible and plausible. Second, the second study further did not reveal any lateralization differences between musicians and non-musicians. For further studies it is necessary to increase statistical power with increased samples of volunteers.

5.3 Advantages of microstate analysis

In the second study of this thesis, current evaluation algorithms were used to evaluate EEG data, namely the microstate analysis. One of the main differences of the microstate analysis lays in the fact that the microstate analysis is independent of a priori selected electrodes or pools of electrodes. The microstate analysis has a global, holistic view to the whole set of data. In the aforementioned study, microstate analysis perfectly examined the N1 and P2 topographies on a holistic level. This study yielded in the duration of the P2 microstate a perfect marker which on the one hand

reflects group differences between musicians and non-musicians in perceptual learning on the other hand reflecting the progression of perceptual learning. Finally, it must be considered that the microstate analysis ist a powerful an novel approach for investigating electrophysiological data in the context of speech perception and musical expertise.

In the light of this facts, the **fourth open question**:

Is it possible from a methodological point of view to show any perceptual learning effect in a microstate analysis?

must be accepted provisionally.

It has been clearly shown that microstate duration can be used as a marker for acoustic expertise, as well as for the evaluation of perceptual learning effects.

5.4 Conclusion

In this thesis, I presented two studies investigating the influence of musical expertise on speech perception in either an active perceptual learning task or an unattended passive listening task. In both studies, musicians showed an advantage in speech perception. I have shown that musical expertise influences both, top-down processes as well as bottom-up processes of speech perception. On a basal layer, musicians are advantaged to extract from auditorilly presented speech signals more information than non-musicians. Otherwise, they also show an advantage on higher levels of cognition in speech processing.

In this thesis, I first used a new method to analyse EEG data in the context of musical expertise and speech, namely the microstate analysis. This new method was a perfect way to investigate the recorded data. It brought to light that the duration of the P2 microstate is an excellent marker to investigate perceptual learning progression.

6 References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci, 28*(15), 3958-3965.

Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J Cogn Neurosci, 22*(2), 392-403.

Aleman, A., Nieuwenstein, M. R., Bocker, K. B., & de Haan, E. H. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*(12), 1664-1668.

Annett, J. (1970). Role of Action Feedback in Acquisition of Simple Motor Responses. *Journal of Motor Behavior, 2*(3), 217-221.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Baumann, S., Meyer, M., & Jancke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J Cogn Neurosci, 20*(12), 2238-2249.

Ben-David, B. M., Campeanu, S., Tremblay, K. L., & Alain, C. (2010). Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology*.

Besson, M., Chobert , J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention and memory. [Review]. *Frontiers in Psychology, 2*.

Besson, M., Schon, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor Neurol Neurosci, 25*(3-4), 399-410.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res, 1355*, 112-125.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience, 33*(3), 530-538.

Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (1999). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology, 20*(4), 615-636.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *The Journal of Neuroscience, 17*(1), 353-362.

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS ONE, 6*(7), e21458.

Brandeis, D., Naylor, H., Halliday, R., Callaway, E., & Yano, L. (1992). Scopolamine effects on visual information processing, attention, and event-related potential map latencies. *Psychophysiology, 29*(3), 315-336.

Brechmann, A., Gaschler-Markefski, B., Sohr, M., Yoneda, K., Kaulisch, T., & Scheich, H. (2007). Working Memory–Specific Activity in Auditory Cortex: Potential Correlates of Sequential Processing and Maintenance. *Cerebral Cortex, 17*(11), 2544-2552.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Can J Psychol, 44*(3), 400-413.

Broca, P. (1861). Remarks on the Seat of the Faculty of Articulated Language, Following an Observation of Aphemia (Loss of Speech). *Bulletin de la Société Anatomique, 6*, 330-357.

Brochard, R., Dufour, A., & Despres, O. (2004). Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery. *Brain Cogn, 54*(2), 103-109.

Brysbaert, M. (2007). "The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem. Royal Holloway: University of London.

Chobert, J., François, C., Habib, M., & Besson, M. (2012). Deficit in the preattentive processing of syllabic duration and VOT in children with dyslexia. *Neuropsychologia, 50*(8), 2044-2055.

Chobert, J., Marie, C., Francois, C., Schon, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *J Cogn Neurosci, 23*(12), 3874-3887.

Clark, H. H. (1973). Language as Fixed-Effect Fallacy - Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Draganova, R., Wollbrink, A., Schulz, M., Okamoto, H., & Pantev, C. (2009). Modulation of auditory evoked responses to spectral and temporal changes by behavioral discrimination training. *BMC Neurosci, 10*, 143.

Eichele, T., Nordby, H., Rimol, L. M., & Hugdahl, K. (2005). Asymmetry of evoked potential latency to speech sounds predicts the ear advantage in dichotic listening. *Cognitive Brain Research, 24*(3), 405-412.

Elmer, S., Meyer, M., & Jancke, L. (2011). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cereb Cortex*.

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex, 22*(3), 650-658.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex, 12*(10), 1031-1039.

Francois, C., Chobert, J., Besson, M., & Schon, D. (2012). Music Training for the Development of Speech Segmentation. *Cereb Cortex*.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain, 129*(Pt 10), 2593-2608.

Gaab, N., Tallal, P., Kim, H., Lakshminarayanan, K., Archie, J. J., Glover, G. H., & Gabrieli, J. D. (2005). Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. *Ann N Y Acad Sci, 1060*, 82-88.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y. X., Hsieh, L., . . . Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*(3), 318-336.

Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception – training increases hierarchical metre processing. *European Journal of Neuroscience, 32*(11), 1979-1985.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron, 56*(6), 1127-1134.

Gordon, E. (1989). *Advanced Measures of Music Audiation*. Chicago:: GIA Publications.

Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbely, A. A., & Achermann, P. (2004). Human central auditory plasticity associated with tone sequence learning. *Learn Mem, 11*(2), 162-171.

Hillyard, S. A. (1981). Selective auditory attention and early event-related potentials: a rejoinder. *Can J Psychol, 35*(2), 159-174.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music Training Improves Verbal but Not Visual Memory: Cross-Sectional and Longitudinal Explorations in Children. *Neuropsychology, 17*(3), 439-450.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). The effects of musical training on structural brain development: a longitudinal study. *Ann N Y Acad Sci, 1169*, 182-186.

Hyde, M. (1997). The N1 response and its applications. *Audiol Neurootol, 2*(5), 281-307.

Jacobsen, T., & Schroger, E. (2003). Measuring duration mismatch negativity. *Clin Neurophysiol, 114*(6), 1133-1143.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric Specialization for Processing Auditory Nonspeech Stimuli. *Cerebral Cortex, 16*(9), 1266-1275.

Jancke, L. (1996). The hand performance test with a modified time limit instruction enables the examination of hand performance asymmetries in adults. *Perceptual and Motor Skills, 82*(3), 735-738.

Jancke, L. (2009). The plastic human brain. *Restor Neurol Neurosci, 27*(5), 521-538.

Jäncke, L. (2012). The relationship between music and language. [Editorial]. *Frontiers in Psychology, 3*.

Jancke, L., Buchanan, T., Lutz, K., Specht, K., Mirzazade, S., & Shah, N. J. (1999). The time course of the BOLD response in the human auditory cortex to acoustic stimuli of different duration. *Brain Res Cogn Brain Res, 8*(2), 117-124.

Jancke, L., Buchanan, T. W., Lutz, K., & Shah, N. J. (2001). Focused and nonfocused attention in verbal and emotional dichotic listening: an FMRI study. *Brain Lang, 78*(3), 349-363.

Jancke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett, 266*(2), 125-128.

Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002a). Phonetic perception and the temporal cortex. *NeuroImage, 15*(4), 733-746.

Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002b). Phonetic perception and the temporal cortex. *NeuroImage, 15*(4), 733-746.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123 (Pt 1)*, 155-163.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Kast, M., Elmer, S., Jancke, L., & Meyer, M. (2010). ERP differences of pre-lexical processing between dyslexic and non-dyslexic children. *International Journal of Psychophysiology, 77*(1), 59-69.

Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport, 10*(6), 1309-1313.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat Rev Neurosci, 11*(8), 599-605.

Kraus, N., Skoe, E., & Parbery-Clark, A. (2008). *Auditory Processing of Pitch, Timbre and Time: Implications for Language and Music*. Paper presented at the 2008 Research Symposium: HEAR OUR VOICES, Milwaukee, Wis.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann N Y Acad Sci, 1169*, 543-557.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language, 110*(3), 135-148.

Krzanowski, W. J., & Lai, Y. T. (1988). A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics, 44*(1), 23-34.

Kühnis, J., Elmer, S., Meyer, M., & Jancke, L. (2012). Musicianship Boosts Perceptual Learning of Pseudoword-Chimeras: An Electrophysiological Approach. *Brain Topogr*.

Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J Neurosci, 26*(15), 4046-4053.

Lackner, J. R., & Teuber, H. L. (1973). Alterations in auditory fusion thresholds after cerebral injury in man. *Neuropsychologia, 11*(4), 409-415.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *J Neurosci, 29*(18), 5832-5840.

Lehmann, D. (1987). Principles of spatial analysis. In A. Gevins & A. Remond (Eds.), *Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 1: Methods of Analysis of Brain Electrical and Magnetic Signals* (pp. 309-354). Amsterdam: Elsevier.

Lehmann, D., & Michel, C. M. (2011). EEG-defined functional microstates as basic building blocks of mental processes. *Clinical Neurophysiology, 122*(6), 1073-1074.

Lehrl, S. (1977). *Mehrfachwahl-WortschatzIntelligenz Test (MWT-B)*. Erlangen: Perimed,.

Lehrl, S., Gallwitz, A., Blaha, V., & Fischer, B. (1991). *Theorie und Messung der geistigen Leistungsfähigkeit mit dem Kurztest KAI.* Ebersberg: Vless.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am, 49*(2), Suppl 2:467+.

Lisker, L., & Abramson, A. S. (1964). Across language study of voicing in initial stops: acoustical measurements. *Word 20*, 384–411.

Liu, C., & Kewley-Port, D. (2004). Vowel formant discrimination for high-fidelity speech. *The Journal of the Acoustical Society of America, 116*(2), 1224-1233.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39*(4), 723-730.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001-1010.

Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J Cogn Neurosci, 18*(2), 199-211.

Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex, 48*(4), 447-457.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *J Cogn Neurosci, 23*(2), 294-305.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19*(9), 1453-1463.

McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends Cogn Sci, 16*(5), 269-276.

Meyer, M. (2008). Functions of the left and right posterior temporal lobes during segmental and suprasegmental speech perception. *Zeitschrift Fur Neuropsychologie, 19*(2), 101-115.

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp, 17*(2), 73-88.

Meyer, M., Baumann, S., & Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *NeuroImage, 32*(4), 1510-1523.

Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restor Neurol Neurosci, 25*(3-4), 411-431.

Meyer, M., Elmer, S., Ringli, M., Oechslin, M. S., Baumann, S., & Jancke, L. (2011a). Long-term exposure to music enhances the sensitivity of the auditory system in children. *Eur J Neurosci, 34*(5), 755-765.

Meyer, M., Elmer, S., Ringli, M., Oechslin, M. S., Baumann, S., & Jancke, L. (2011b). Long-term exposure to music enhances the sensitivity of the auditory system in children. *European Journal of Neuroscience, 34*(5), 755-765.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang, 89*(2), 277-289.

Meyer, M., Zaehle, T., Gountouna, V. E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport, 16*(18), 1985-1989.

Michel, C. M. (2009a). *Electrical neuroimaging*. Cambridge: Cambridge University Press.

Michel, C. M., Koenig, Th., Brandeis, D., Gianotti, L. R. R., and Wackermann, J. (2009b). Electrical neuroimaging in the time domain - microstate analysis. In M. C. M., Koenig, Th., D. Brandeis, L. R. R. Gianotti, and J. Wackermann (Ed.), *Electrical Neuroimaging* (pp. 123-125). Cambridge: Cambridge University Press.

Michel, C. M., Seeck, M., & Landis, T. (1999). Spatiotemporal Dynamics of Human Cognition. *News Physiol Sci, 14*, 206-214.

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: a review. [Review]. *Frontiers in Psychology, 2*.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol Sci, 22*(11), 1425-1433.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cereb. Cortex, 19*(3), 712-723.

Mulert, C., Jager, L., Propp, S., Karch, S., Stormann, S., Pogarell, O., . . . Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI. *NeuroImage, 28*(1), 49-58.

Mulert, C., Jager, L., Schmitt, R., Bussfeld, P., Pogarell, O., Moller, H. J., . . . Hegerl, U. (2004). Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage, 22*(1), 83-94.

Munte, T. F., Altenmuller, E., & Jancke, L. (2002). The musician's brain as a model of neuroplasticity. [10.1038/nrn843]. *Nat Rev Neurosci, 3*(6), 473-478.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr, 20*(4), 249-264.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage, 21*(1), 125-135.

Murray, M. M., Wylie, G. R., Higgins, B. A., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *J Neurosci, 22*(12), 5055-5073.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NY: Lowrence Erlbaum.

Naatanen, R., & Alho, K. (1995). Mismatch negativity--a unique measure of sensory processing in audition. *Int J Neurosci, 80*(1-4), 317-337.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313-329.

Naatanen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biol Psychol, 8*(2), 81-136.

Naatanen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clin Neurophysiol, 115*(1), 140-144.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear, 30*(4), 432-446.

Oechslin, M. S., Meyer, M., & Jancke, L. (2010). Absolute pitch--functional evidence of speech-relevant auditory acuity. *Cereb Cortex, 20*(2), 447-455.

Oechslin, M. S., Van De Ville, D., Lazeyras, F., Hauert, C. A., & James, C. E. (2012). Degree of Musical Expertise Modulates Higher Order Brain Functioning. *Cereb Cortex*.

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear and Hearing, 19*(4), 290-297.

Ott, C. G., Langer, N., Oechslin, M., Meyer, M., & Jancke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Front Psychol, 2*, 195.

Pantev, C., Engelien, A., Candia, V., & Elbert, T. (2001a). Representational cortex in musicians. Plastic alterations in response to musical practice. *Ann N Y Acad Sci, 930*, 300-314.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811-814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001b). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport, 12*(1), 169-174.

Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*(0), 111-119.

Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. (2011). Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles. *Psychophysiology*.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology, 24*, 91-95.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low-Resolution Electromagnetic Tomography - a New Method for Localizing Electrical-Activity in the Brain. *International Journal of Psychophysiology, 18*(1), 49-65.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans Biomed Eng, 42*(7), 658-665.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. [Hypothesis & Theory]. *Frontiers in Psychology, 2*.

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of Scalp Potentials by Surface Spline Interpolation. *Electroencephalography and Clinical Neurophysiology, 66*(1), 75-81.

Phillips, C., Rugg, M. D., & Friston, K. J. (2002). Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. *NeuroImage, 16*(3 Pt 1), 678-695.

Phillips, D. P., & Hall, S. E. (2002). Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra. *Hear Res, 174*(1-2), 133-141.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., & Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: Some properties of ''central'' auditory gap detection. *Journal of the Acoustical Society of America, 101*(6), 3694-3705.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr Clin Neurophysiol, 36*(2), 191-199.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. [doi: DOI: 10.1016/S0167-6393(02)00107-3]. *Speech Communication, 41*(1), 245-255.

Pratt, H., Starr, A., Michalewski, H. J., Bleich, N., & Mittelman, N. (2007). The N1 complex to gaps in noise: effects of preceding noise duration and intensity. *Clin Neurophysiol, 118*(5), 1078-1087.

Rajan, R., Irvine, D. R., Wise, L. Z., & Heil, P. (1993). Effect of unilateral partial cochlear lesions in adult cats on the representation of lesioned and unlesioned cochleas in primary auditory cortex. *J Comp Neurol, 338*(1), 17-49.

Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci, 18*(9), 383-388.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci, 13*(1), 87-103.

Russeler, J., Altenmuller, E., Nager, W., Kohlmetz, C., & Munte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci Lett, 308*(1), 33-36.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & Ohl, F. W. (2007). The cognitive auditory cortex: Task-specificity of stimulus representations. *Hearing Research, 229*(1–2), 213-224.

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann N Y Acad Sci, 930*, 281-299.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*(5198), 699-701.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Ann N Y Acad Sci, 1060*, 219-230.

Schlegel, F., Lehmann, D., Faber, P. L., Milz, P., & Gianotti, L. R. (2011). EEG Microstates During Resting Represent Personality Differences. *Brain Topogr*.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005a). Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann N Y Acad Sci, 1060*, 387-394.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., . . . Rupp, A. (2005b). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci, 8*(9), 1241-1247.

Schön, D., & Francois, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. [Review]. *Frontiers in Psychology, 2*.

Schon, D., Gordon, R. L., & Besson, M. (2005). Musical and linguistic processing in song perception. *Neurosciences and Music Ii: From Perception to Performance, 1060*, 71-81.

Schon, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

Schreiner, C. E., & Winer, J. A. (2007). Auditory Cortex Mapmaking: Principles, Projections, and Plasticity. *Neuron, 56*(2), 356-365.

Schwent, V. L., & Hillyard, S. A. (1975). Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroencephalogr Clin Neurophysiol, 38*(2), 131-138.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci, 23*(13), 5545-5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport, 16*(16), 1781-1785.

Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Front Psychol, 2*, 126.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America, 108*(6), 3030-3035.

Shen, C., Smith, Z. M., Oxenham, A. J. & Delgutte, B. (2001). Auditory Chimera Demo, from http://epl.meei.harvard.edu/~bard/chimeral

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *Journal of Neuroscience, 27*(14), 3799-3806.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*(6876), 87-90.

Studdert-Kennedy, M., & Shankweiler, D. (1981). Hemispheric specialization for language processes. *Science, 211*(4485), 960-961.

Sturm, W., & Willmes, K. (1999). *Verbaler Lerntest und Nonverbaler Lerntest (VLT / NVLT)*. Göttingen: Hogrefe.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp Brain Res, 161*(1), 1-10.

Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur J Neurosci, 30*(8), 1636-1642.

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics, 14*(3), 511-528.

Trainor, L. J., Shahin, A., & Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Ann N Y Acad Sci, 999*, 506-513.

Tzounopoulos, T., & Kraus, N. (2009). Learning to Encode Timing: Mechanisms of Plasticity in the Auditory Brainstem. *Neuron, 62*(4), 463-469.

van Zuijen, T. L., Sussman, E., Winkler, I., Naatanen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity--a mismatch negativity study comparing musicians and non-musicians. *Brain Res Cogn Brain Res, 23*(2-3), 270-276.

Vaughan, H. G., Jr., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalogr Clin Neurophysiol, 28*(4), 360-367.

Vigneau, M., Beaucousin, V., Herve, P. Y., Jobard, G., Petit, L., Crivello, F., . . . Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*(1), 577-593.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music--brain responses to rhythmic incongruity. *Cortex, 45*(1), 80-92.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Ostergaard, L. (2005). To musicians, the message is in the meter - Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage, 24*(2), 560-564.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*.

Wetzel, W., Ohl, F. W., & Scheich, H. (2008). Global versus local processing of frequency-modulated tones in gerbils: an animal model of lateralized auditory cortex functions. *Proc Natl Acad Sci U S A, 105*(18), 6753-6758.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci, 10*(4), 420-422.

Zaehle, T., Jancke, L., Herrmann, C. S., & Meyer, M. (2009). Pre-attentive spectro-temporal feature processing in the human auditory system. *Brain Topogr, 22*(2), 97-108.

Zaehle, T., Jancke, L., & Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav Brain Funct, 3*, 63.

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cereb. Cortex, 11*(10), 946-953.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 1087-1104.